DOCUMENT RESUME

ED 423 892 IR 057 194

AUTHOR	Gonzalez, Pedro
TITLE	Computerization of the Archivo General de Indias: Strategies and Results.
INSTITUTION	Council on Library and Information Resources, Washington, DC.
ISBN	ISBN-1-887334-61-0
PUB DATE	1998-09-00
NOTE	66p.
AVAILABLE FROM	Council on Library and Information Resources, 1755 Massachusetts Avenue, N.W., Suite 500, Washington, DC 20036 ($20).
PUB TYPE	Reports - Descriptive (141)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	*Archives; Case Studies; Decision Making; Electronic Libraries; Foreign Countries; Information Storage; Information Systems; Library Administration; *Library Automation; Library Collection Development; *Library Materials; Reference Materials
IDENTIFIERS	*Digital Data; *Digital Technology; Spain

ABSTRACT

The project of the Archivo General de Indias (AGI) in Seville, Spain has digitized more than 11 million of its pages of documents relating to Spanish history in the New World. This report illustrates the experience of the AGI, and the range of decisions that managers have faced throughout the project. Following an introduction and background information, the report focuses on the following: project objectives of Stage I (1986-1992) and Stage II (1993-present); general system architecture; information and reference subsystem; digital image storage system; technical aspects; and new prospects for long-distance access. In a concluding section, the multidirectional progress, project costs and results, and final observations are outlined. Appendices include an analysis of conservation results, analysis of consultation results, analysis of results of "supplemental" digitization of documents from other archives, and figures for costs and production. A bibliography is also included. (AEF)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 423 892

Computerization of the Archivo General de Indias: Strategies and Results

by Pedro González

September 1998

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
B.H. Deney
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Council on Library and Information Resources

Commission on Preservation and Access
Digital Libraries
Economics of Information
Leadership

IR057194

BEST COPY AVAILABLE

2

About the Author

Pedro González is an archivist with many years' experience in the application of information technologies to historical documents. From 1987 to 1992, as director of the Centro de Información Documental de Archivos, he represented the Ministry of Culture on the team directing the Archivo General de Indias' computerization project. He then served as director of the Archivo from 1993 to 1997 and during his tenure was responsible for the project's operational phase.

The author has served on several international committees on archives, including the Experts Group on Archival Coordination at the European Union, the Automation and Image Technology committees of the International Council on Archives (ICA), and the ICA's Ad Hoc Commission for Descriptive Standards.

Commission on Preservation and Access

The Commission on Preservation and Access, a program of the Council on Library and Information Resources, supports the efforts of libraries and archives to save endangered portions of their paper-based collections and to meet the new preservation challenges of the digital environment. Working with institutions around the world, the Commission disseminates knowledge of best preservation practices and promotes a coordinated approach to preservation activity.

Digital Libraries

The Digital Libraries program of the Council on Library and Information Resources is committed to helping libraries of all types and sizes understand the far-reaching implications of digitization. To that end, CLIR supports projects and publications whose purpose is to build confidence in, and increase understanding of, the digital component that libraries are now adding to their traditional print holdings.

ISBN 1-887334-61-0

Published by:

Council on Library and Information Resources
1755 Massachusetts Avenue, N.W., Suite 500
Washington, D.C. 20036

Additional copies are available for $20.00 from the above address. Orders must be prepaid, with checks made payable to the Council on Library and Information Resources.

The paper in this publication meets the minimum requirements of the American National Standard for Information Sciences -Permanence of Paper for Printed Library Materials ANSI Z39.48-1984.

Copyright 1998 by the Council on Library and Information Resources. No part of this publication may be reproduced or transcribed in any form without permission of the publisher. Requests for reproduction for noncommercial purposes, including educational advancement, private study, or research will be granted. Full credit must be given to both the author and the Council on Library and Information Resources.

Computerization of the Archivo General de Indias: Strategies and Results

by Pedro González

Council on Library and Information Resources
Washington, D.C.

September 1998

CONTENTS

FOREWORD v

AUTHOR'S ACKNOWLEDGMENTS vi

INTRODUCTION 1

- Lessons of a Ten-Year Experience 1
- Range of Achievements 2
- Project Origin 3
- Project Characteristics 3
- Project Challenges 4
- Importance of Results for Conservation and Access 4
- The Archivo General de Indias and the "Global Village" 4

BACKGROUND 5

- Founding 5
- Documentary Holdings 5
- Organization of Holdings 5
- Finding Aids 6
- The AGI Building 6
- Consultation of Holdings 7
- Restoration Laboratory 7
- Microfilm Service 7

PROJECT OBJECTIVES 8

- Stage I: Objectives 1986-1992 8
- Stage II: Objectives 1993-present 9

GENERAL SYSTEM ARCHITECTURE 10

- Fundamental System Modules 10
- General Architecture 11

INFORMATION AND REFERENCE SUBSYSTEM **12**

Description at the AGI 12
Finding Aids 13
Objectives for the Information and Reference System 15
Outline of Information and Reference System 15
Data Access by the "Principle of Provenance Path" 15
Direct Information Access 16
AGI Indices 16
Keywords 17
Elements of Description 19
Basic Information Area 19
Descriptive Information Area 19
Elements of Location, Retrieval, or Access 20
Initial Content: Retrospective Conversion of Finding Aids 20
The Information and Reference System Today 21
Direct and Indirect Access to Archival Information 22
Information Retrieval Tools 22
Standardization 23
Revision and Updating of Information System Content 23

DIGITAL IMAGE STORAGE SYSTEM **24**

Digitization Process 24
Selection of Documents to be Digitized 24
Document Preparation 25
Digitization 26
Image Quality 26
Color Images 28
Storage Medium 29
Formats and Image Compression 30
Quality Control 31
Image Consultation 31
Disk Service 31
Image Consultation 33

TECHNICAL ASPECTS **34**

Initial Computer Equipment 34
Current Equipment 35
Other Current Equipment Possibilities 35

NEW PROSPECTS: LONG-DISTANCE ACCESS **37**

Technical Problems 38
Security Problems 38
Problems of Intellectual Property and Document Ownership 38

Management Problems 39

CONCLUSIONS **40**

Multidirectional Progress 40
Project Costs 40
Project Results 42
Conservation Benefits 42
Access Benefits 42
Benefits for Internal Operation 43
Final Observations 43

GENERAL BIBLIOGRAPHY **44**

APPENDIX I: Analysis of Conservation Results 47

APPENDIX II: Analysis of Consultation Results 49

APPENDIX III: "Supplemental" Digitization of Documents from Other Archives: Analysis of Results 53

APPENDIX IV: Figures for Costs and Production 55

FOREWORD

Today, many archives and libraries around the world have embarked upon projects to digitize portions of their holdings. Motivations for this work vary, but often the goal is to broaden access, facilitate retrieval, and reduce handling of the originals.

With rapidly changing technologies and costs, planning for the management of such projects over time becomes difficult. As a result, much digitizing work today is undertaken in "pilot projects" that allow institutions to learn by doing and learn from each other.

The project of the Archivo General de Indias (AGI) in Seville, Spain, is an important case study both for its size and for its track record. Over the past decade, the AGI has digitized more than eleven million pages of documents relating to Spanish history in the New World. In addition, its system for providing access to the digital documents has been in use for five years. The Archives has thus had to deal with the day-to-day practical problems of operation and longer-term issues such as obsolescence of hardware, software, and storage media.

From the start, the AGI was clear about its rationale for digitizing to "preserve." Its objective was not to create a preservation copy in digital form to replace originals, but to offer digital surrogates to reduce the handling of originals. Today, about one-third of AGI's on-site consultations are done electronically, greatly reducing exposure of original documents.

The report illustrates the range of difficult decisions that managers have faced throughout the project. Decisions often had to be made where no precedent existed. And, as always, decisions were bound by time and money, forcing choices that were not always optimal, but realistic.

We hope that the experience of the AGI, as described in the following report, will be a useful case study for planners facing the myriad technical, organizational, and managerial challenges presented by their own digitizing projects.

AUTHOR'S ACKNOWLEDGMENTS

The author wishes to acknowledge all the people at the Archivo General de Indias who have collaborated in the design, development, installation, and operation of the computerized system. The Archivos y Bibliotecas group, from Informática El Corte Inglés, gave its close collaboration throughout the project and its advice in preparing the work.

In the United States, several people collaborated in bringing this report to publication. Ruth Morales translated the original work from Spanish. Dan Hazen and Charles Faulhaber read through the paper, and Kathlin Smith edited the English version. Hans Rütimann, long interested in the project, encouraged the writing and publication of this report.

INTRODUCTION

Over the past decade, the large-scale digital reformatting project of the Archivo General de Indias (AGI) in Seville has received considerable attention for its application of computer technologies to historical archives. In 1991, Hans Rütimann and Stuart Lynn prepared a report on the project for the Commission on Preservation and Access. The report analyzed advantages and problems and concluded that the AGI project was a useful model for other projects: "There is much to learn for other applications."[1] Jean-Paul Oddos published a later report on the project in the *Bulletin de la Bibliothèque de France* in 1994. It contains a positive assessment in conclusions entitled "Un démarche exemplaire."[2] In a UNESCO RAMP report, Christopher Kitching commented on the profession's interest in the project's outcome: "Its results will be eagerly awaited by the archival world."[3] At the official opening of the system, presided over by the King and Queen of Spain, Charles Kesckeméti, Secretary General of the International Council on Archives, began his remarks as follows: "This is a fantastic day in the history of archives."

These are just a few statements from the many professionals who have visited Seville to study the project. UNESCO itself supported the organization of two information workshops on the project for Latin American professionals. The International Council on Archives, backed by the Council of Europe, has promoted the Seville model in computerizing the Komintern archives in Moscow.

Lessons of a Ten-Year Experience

It has been 11 years since the project was launched by an agreement between the Ministry of Culture (under which the AGI operates), the Ramón Areces Foundation, and IBM España. This is a very long stretch in the changing world of new technologies, where everything quickly becomes dated and obsolete. It is enough time for us to as-

[1] Hans Rütimann and Stuart Lynn, *Computerization Project of the Archivo General de Indias, Seville, Spain. A Report to the Commission on Preservation and Access* (Washington, DC: Commission on Preservation and Access, 1992).

[2] Jean-Paul Oddos, "La Numérisation des Archives Générales des Indes à Seville" [Digitization of the Archivo General de Indias in Seville], *Bulletin de la Bibliothèque de France* 39(4) (1994): 82-86.

[3] Christopher Kitching, *The Impact of Computerization on Archival Finding Aids: A RAMP Study* (Paris: UNESCO, 1991).

sess, with some perspective, a project that has been the most important ongoing effort in applying new digital technologies to the treatment of historical archives.[4] There is an additional advantage in undertaking this analysis: because of its continuity, all phases of the project can be analyzed, from conception to operation and consolidation. Over those nearly 11 years, the system was designed, developed, and installed at the AGI. Today, the system is a fundamental part of AGI's operations and its use has been established among researchers and staff. In addition, many steps have been taken to update the system and solve the problems posed by rapid technological obsolescence.

Range of Achievements

The AGI computerization project has made several advances in the application of new technologies to archives. No historical archive in the world today can claim a greater variety of achievements:

- Eleven million pages are available for direct on-screen consultation, supported by a spectacular combination of image-enhancing tools.
- More than 30 percent of consultations at the AGI are done electronically.
- All descriptive information is available electronically, forming a unified data system into which all traditional finding aids have been integrated following their retrospective conversion.
- This entire integrated system is in daily operation in the Reading Room.

Initiation of the project required identification, discussion, and analysis of the many problems and the possibilities offered by new technologies. Only then was it possible to develop a complete system for managing historical archives, with special emphasis on the latest aspects of image treatment. But the most remarkable aspect of the project is that the system has been functioning for almost five years. During this time, staff have had to address the day-to-day practical problems of operation, including staff training, aging equipment, and, above all, obsolescence of hardware, software, and data media. The AGI is a laboratory where an integrated computerized system responsible for most of the functions of a historical archive can be observed operating in real time.

[4] The agreement establishing the AGI computerization project was signed in 1986, effective to December 1992. Textual data entry began in 1988 and digitization of documents in 1989. The first version of the user-management module was installed in mid-1988. The complete system was installed on a trial basis in 1991 and officially inaugurated on October 6, 1992. At the conclusion of the initial project, a new agreement was signed covering 1993 and 1994.

Project Origin

To understand the project's aims, methodology, and results, it is worth noting the circumstances leading to its inception. Why was such an ambitious and massive project undertaken? Why is this project, unique in the often-neglected field of treatment of historical documentation, being carried out in Spain?

The main impetus for the project was Spain's celebration of the fifth centenary of the discovery of America, which culminated in the World's Fair in Seville in 1992. For several years before the celebration, cultural, entrepreneurial, and political agencies had been encouraged to consider ways to take advantage of the possibilities offered by such a significant public event.

The Archivo General de Indias offered an excellent stage for activities relating to the commemoration. For more than a century, researchers from all over the world have been drawn by its stacks of documents of basic interest for studying the discovery and the history of the Americas. Accordingly, in 1986, three organizations agreed to carry out the computerization project at the Archivo. The organizations included a public institution (the Ministry of Culture, under which the AGI operates), a nonprofit private cultural institution (Ramón Areces Foundation), and a computer technology company (IBM España). This public-private partnership was critical for securing the heavy investments needed by the project and gave all participants increased visibility for their support of an important cultural activity.

Project Characteristics

- The project was designed to apply the latest computer technologies, especially digital imaging, to the treatment of historical archives in support of the AGI's goals of preservation and access.
- The project set specific dates for completion and the reporting of results, requiring that by 1992 a system be installed and operating. Thus, the system already incorporated a large volume of data when it opened, rather than appearing as an empty container to be filled.
- The project was aided by the participants' commitment to support cultural activities. This meant that financial resources were allocated for project development and that staff from the three institutions would collaborate closely over several years.
- The project was intended to serve as the pilot project for the eventual computerization of all Spanish national historical archives.

Project Challenges

- The work involved the use of new, developing technology, especially in image treatment and information medium. Challenges included the rapidity of technical developments, lack of standards, high cost, and risk of obsolescence.
- The AGI had no experience in the use of the new imaging technologies.
- The project required large investments because of the high cost of new technologies. (The costs did, however, decrease considerably over time.)
- Certain technological decisions had to be put off as long as possible to present a state-of-the-art system in 1992.
- The project had to progress simultaneously in system design and development and in massive data entry, so that by 1992 the project would be operating at the AGI and a large volume of digital information would already be available.
- The project had to remain within the scope of what was practical, given the available resources.

Importance of Results for Conservation and Access

AGI now provides more than 30 percent of document consultation service through digital images, eliminating handling of the original and offering significant advantages for both access and conservation. The average time that researchers need to conduct their work is much less than before. There are more researchers, but they complete their work sooner, and therefore need to spend fewer days working in the Archivo.

The Archivo General de Indias and the "Global Village"

The future incorporation of digital holdings into communication networks (which also requires careful analysis) could make the Archivo General de Indias the first historical archive in the world to offer the researcher such a large collection of material by remote access. The emerging "global village" makes this desirable, but its impact on small institutions should be carefully studied to ensure appropriate aims, means, and results.

BACKGROUND[5]

Founding

In 1781, King Carlos III decided to found the AGI at the urging of his Secretary of the Indies, José de Gálvez. The goal was to gather at one site all the documents concerning Spanish administration in the Americas and Philippines. The original documents were to be used as source material for writing a new history of the Spanish presence in the Americas. The new history would respond to some of the histories being published abroad that Spanish officials and intellectuals feared were auguring a new episode of the anti-Spanish "black legend."[6] The move would also free up space at the Simancas Archivo, the central archive of the Crown since the sixteenth century, which was then so overwhelmed it could not offer proper service.

Documentary Holdings

The first shipments of papers from Simancas arrived in October 1785, forming the initial core of the AGI. They would later be supplemented by new transfers, especially from Madrid and Cádiz. Today, the AGI holds about eight kilometers of shelving containing more than 43,000 bundles of original papers.

These documentary holdings are drawn from the metropolitan agencies responsible for colonial administration, primarily the Consejo de Indias (Council of the Indies), Casa de la Contratación (House of Trade), State Secretariats responsible for Indies affairs, and the Consulados (Boards of Trade) of Seville and Cádiz, as well as other agencies.

Organization of Holdings

The AGI was provided with an excellent series of Ordenanzas (ordinances) promulgated by King Carlos IV in 1790.[7] Among other functions, these Ordenanzas established what eventually came to be known as the "principle of provenance." This principle obligated the

[5] The General Bibliography lists useful titles about the AGI's history and holdings.

[6] According to Juan Bautista Muñoz, founder of the Archivo: "In order to fulfill these worthy purposes, to silence once and for all our many fiery defamers and rivals and to show their ignorance to be inexcusable, it was necessary to go to the root of the matter, to the sources, and study irrefutable documents, as if nothing [else] had been written and published." Juan Bautista Muñoz to the Secretary of State of the Indies, José de Gálvez, 28 November 1783. Archivo Histórico Nacional, Diversos 29, Doc. 16

[7] *Ordenanzas para el Archivo General de Indias* [Ordinances for the AGI] Article V (Madrid, 1990). See bibliography on the 1790 Archivo Ordinances in the General Bibliography.

Archives to keep together all the documents generated by each agency,[8] without mixing them with the documents of other agencies.[9] Article V of the Ordenanzas reads: "The first division of papers should be into collections corresponding to the remitting offices. Thus, those from Simancas, Vía Reservada, and each of the offices of the Consejo should remain together and be maintained separately from the others."[10] Although this obligation was observed for the most part, trends in the history of the AGI eventually led to the current organization of documents into 15 sections that usually, but not always, hold all the documents of a unique generating agency. More detail on this organization is given on p. 12.

Finding Aids

During the first few years of its history, the AGI made a great effort to organize and describe its holdings in order to create a "general inventory."[11] Although this general inventory was never completed, the AGI today has many guides, inventories, catalogs, and indices that make it possible to control the holdings and facilitate access to the information. Some of the old inventories, painstakingly drawn up at the end of the eighteenth century, have continued to be useful in their manuscript format.

The AGI Building

The AGI is located in the old building of the Casa Lonja de Mercaderes (Commodity Exchange) of Seville, constructed between 1583 and 1646 as a meeting place for dealers who traded between the metropolis and its colonies. The building was renovated to serve as the headquarters of the AGI at the time it was founded. Besides being a Spanish historical monument,[12] the site has been declared part of the World Heritage by UNESCO.

[8] This principle identifies the "fonds," an archival concept widely used in Europe. The ISAD(G) standard defines the term as the "whole of the documents, regardless of form or medium, organically created and/or accumulated and used by a particular person, family, or corporate body in the course of that creator's activities and functions."

[9] The "principle of provenance" is understood today to include not only the separation of documents generated by each agency but also the conservation of their original order. French archivist Natalis de Wailly first enunciated the principle in 1841 as "respect de fonds." The Germans then developed the Strukturprinzip, which later became part of the "principle of provenance." The Ordenanzas clearly expressed this principle in 1790, even though the terms had not yet been developed.

[10] *Ordenanzas*, Article V.

[11] *Ordenanzas*, Article XXVII.

[12] The site was declared a national monument by Royal Decree on April 20, 1983, and a World Heritage site by UNESCO in 1987.

Consultation of Holdings

The AGI is visited daily by an average of 50 researchers. More than 900 different researchers visit each year. Half come from outside Spain; almost 40 percent of all researchers come from the Americas. The AGI fills requests for 300,000 to 400,000 copies on paper and microfilm each year and responds to almost a thousand written requests annually.

Restoration Laboratory

The AGI has a restoration laboratory to handle its conservation problems. All papers are more than a century old, and some are 500 years old, with a concentration in the sixteenth and seventeenth centuries. The documents have been affected by different storage conditions. Many have also undergone the vicissitudes of a long Atlantic crossing, although great effort was made to ensure optimal packing for shipment. In some cases, the papers and ink themselves have caused degradation. For example, the Philippines Audiencia often used fragile rice paper for documents. In recent years, documents have been excessively handled in the Reading Room.

Microfilm Service

The AGI also contains a small microfilm laboratory which, because of staffing limitations, generally only fulfills researchers' requests for copies. However, a policy of backup microfilming and microfilm editions has gradually been adopted over the years, so that about two million frames in unperforated 35 mm rolls are currently available.[13]

[13] The list of AGI documents available on microfilm through the Document Reproduction Service of the Ministry of Education and Culture is contained in the *Boletín de Información del CIDA* (Centro de Información Documental de Archivos), n. 1 (1993).

PROJECT OBJECTIVES

The goal of the project was to design, develop, and implement in the AGI an automated data system capable of integrated management of the most common functions of a historical archive. The system was intended to offer solutions to the access/conservation dichotomy that is a central concern for any archive.

The goal of producing the automated data system was to be met through a series of more limited objectives, which can be divided into two major stages. The first, extending from 1986 through 1992, involved design, development, massive data entry, and system installation. The second, beginning in 1993, was the phase of actual operation, staff training, consolidation, solution of obsolescence problems, and progress toward new standards.

Stage I: Objectives 1986-1992

- To design and develop the system according to the Archivo's needs, using the latest technologies in databases and digital imaging.
- To undertake massive data entry, so that by 1992 the AGI would have in use not only a system but also abundant information, including both the descriptive information necessary for document location and the digital images.
- To retrospectively convert descriptive information derived from the AGI's inventories, catalogs, and indices. Such conversion also had to incorporate the more detailed descriptions that were compiled for the documents to be digitized.
- To scan for digital images 10 percent of the AGI's holdings, about eight million pages.
- To implement the system once it was developed and put it into service.
- To introduce some parts of the system into other archives, specifically the user-management module, which became the first byproduct for use elsewhere.

In 1992, the initial objectives were met, the system was installed at the AGI, and a new phase of consolidation, expansion, and updating was begun. Although a two-year extension of the agreement among the sponsors made it possible to continue and improve upon what had already been done, the AGI's objectives changed. It now had the job of assuming direct operation of the system: integrating the system into daily operations, training staff and bringing on technicians, finding new sources of financial support, and preparing strategies for equipment update and data migration. This was to be done independently, without the external oversight of the sponsors.

Stage II: Objectives 1993-present

- To make the system operational in the AGI, integrating the daily routines and functions, including reference service, Reading Room management, and documents service. The new tasks of digitization and entry of textual data would also be part of the daily work, along with other common duties such as restoration and microfilming.
- To become independent of external support from the institutions collaborating in the project, and to obtain the necessary resources and make changes in staffing, such as bringing in qualified new personnel and retraining existing staff.
- To review and solve problems caused by rapid technological obsolescence, taking the steps needed to adapt the system to new computer advances.
- To progress toward more open and standard formulas (some of which did not exist at system startup), so that the future exchange of data and the migration to new generations of hardware and software would be easier.
- To address the lack of backup copies of optical disks. This problem had been put aside until inexpensive, effective alternatives appeared in the market.
- To continue and, if possible, expand the work of generating descriptive information and digital images of documents.
- To begin to locate the system within the current framework of communication networks. When the project was started, the possibility of remote access was analyzed, but use of the Internet was still very limited.

The following sections provide detail on the system and subsystems that were developed to serve the AGI.

GENERAL SYSTEM ARCHITECTURE

The system was designed in modular form, based on the idea of distributed processing and client/server architecture. This was done to ensure flexibility for future use of the system in other larger or smaller archives, and took into consideration future access options through networks.

Fundamental System Modules

From the user's standpoint, the system consists of three modules, each of which covers a different group of functions:

The information and reference system provides descriptive information that gives the user all of the tools needed to find original or reproduced documents. It is not simply an index or database providing access to digitized information. It is a unified, global system of information valid for all descriptive information in the Archivo. It includes data contained in all the inventories, catalogs, and indices

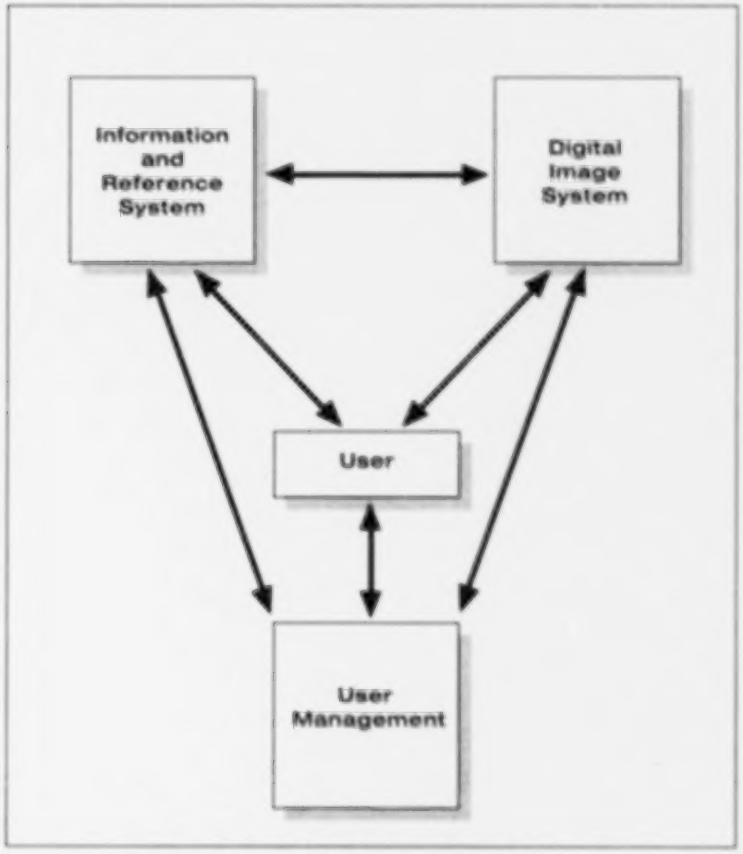

Figure 1. *General Scheme of System Operation*

produced over the AGI's history. It also includes the information necessary to recover digital images of the documents.

The digital image storage system offers the tools needed to consult digital copies on the monitor or on paper. The system handles two independent processes: document digitization and its storage on optical disk, and access to digitized images for consultation.

User management includes all of the functions relating to the retrieval of AGI documentation, through both manual and automated consultation:

- Tracking research requests (researchers' files, topics of investigation, statistics on use, and user profiles)
- Monitoring Reading Room use
- Controlling document movement within the Archivo
- Controlling reproduction service for researchers

General Architecture

To handle the series of functions outlined above and to meet the aims for flexibility, the system consists of various subsystems. Each subsystem comprises a series of logically related functions that are carried out within their corresponding operating environments. Different servers provide the main functions in communication with the several workstations:

These subsystems are as follows:

- **The textual database** deals with operation of the information and reference system. It includes the data entry and consultation processes.
- **Digital image storage** provides the resources for consultation of digital surrogates. It includes the processes of digitization and consultation.
- **The user-management database** permits control of the operation of archive and system consultation.
- **The liaison interface** provides the necessary communication with the other subsystems.
- **The user interface** facilitates user dialogue at the "client" station. It converts user inquiries into messages sent to the corresponding subsystem, and converts the messages received into easily intelligible replies.
- **The printing subsystem** allows for printing textual information and digital images.

INFORMATION AND REFERENCE SYSTEM

The information reference subsystem, which provides access to descriptive information and thereby to the content of the Archivo's original documents, is the heart of the entire system. It was designed to meet the objectives outlined on p. 15.

Description at the AGI

The AGI is a historical archive organized according to traditional archival principles, which differentiate the archive from the library or the documentation center. Its organization is further specified in the "principle of provenance" noted on pp. 5-6. Most of the AGI is organized according to this principle of provenance. The first division consists of sections that usually contain the papers of a single producing agency or unit thereof. The sections may be divided into subsections (especially Section V, Government, which is organized into subsections according to the territorial organization of the several Audiencias and Viceroyalties). The sections in turn are broken down into series (sometimes redivided into subseries) containing those documents produced by the corresponding agency in the exercise of each of its functions. The series are made up of documents or files, grouped into bundles.

The resulting hierarchical and multilevel model of organization is shown below:

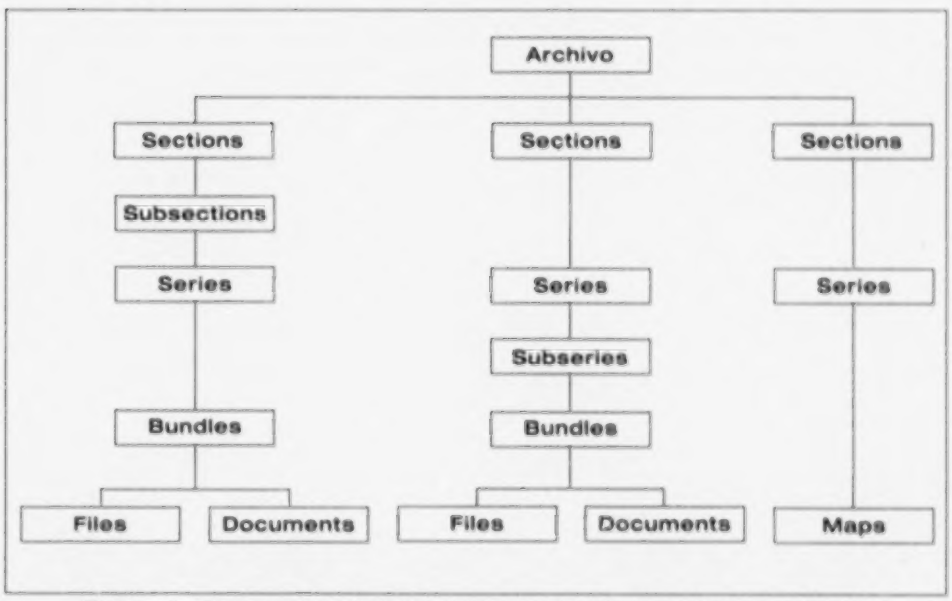

Figure 2. *Organization Model of Holdings in the Archivo General de Indias*

Finding Aids

The 1790 Ordenanzas also mandated the formation of various finding aids that would provide physical and intellectual control of the AGI's holdings and the necessary lines of access to the information. The ultimate objective was to compile a complete and systematic "general inventory" of all the organized holdings of the Archivo.[14] This work program prefigured to some extent what we now call an "information system."

In practice, the archival description consists of individual finding aids, such as guides, inventories, catalogs, and indices. Although the systematic "general inventory" mentioned in the AGI Ordenanzas is not complete, these independent instruments aim to project a complete and systematic vision consistent with the hierarchical structure of the holdings.

The characteristics of archival description are as follows:

- It is multilevel, including descriptions or entries of various archival units or groupings considered on their several levels: collection, series, subseries, file, and document.
- It reflects the hierarchical family-tree structure of archive organization, in which archival units or groupings are included in others at higher levels and may in turn include others at a lower level.
- For the description of each document or piece to be wholly meaningful and fully intelligible, it must relate to the higher-level units of which it is a part and, ultimately, to the agency producing the documents.
- The principle of provenance thus becomes not only the basic norm for document organization but also the essential element for attaining intellectual control of the holdings and for guiding information access.[15]
- This indirect access to information is supplemented by indices leading directly to the subject, person, and place. But the indices are generally auxiliary or complementary to the principal finding aids. This means that not even access through indices is totally direct; the inventory or catalog must be located first to use its auxiliary index.

Accordingly, over the two centuries of AGI history, numerous finding aids have been developed (and occasionally published): a

[14] *Ordenanzas*, Article XXVII.

[15] See Eric Ketelaar, "Exploitation of New Archival Materials," in *Proceedings of the 11th International Congress on Archives* (Munich, New York, London, Paris: K.G. Sauer, 1989), 189-99; Richard H. Lytle, "Intellectual Access to Archives," *American Archivist* 43 (1980): 64-76 and 191-208; David A. Bearman and Richard H. Lytle, "The Power of the Principle of Provenance," *Archivaria* 21 (1985-86): 14-27; and Michel Duchein, "Theoretical Principles and Practical Problems of Respect des Fonds," *Archivaria* 16 (Summer 1983): 64-82.

BEST COPY AVAILABLE

general guide to the Archivo,[16] inventories for each section,[17] catalogs of certain subjects or series,[18] and indices.

Many changes in personnel, goals, and criteria relating to the work have taken place since the AGI was founded. Consequently, there are great differences among the various finding aids—in format, depth of description, level of detail, and terminology.

There are manuscript instruments, such as inventories for the Sections of Patronato (relations of the Kings with the Church), Contaduría (General Accounting), Contratación (House of Trade or House of the Indies), Justicia (Justice); typewritten texts, such as the inventories for the Sections of Gobierno (Government, the papers of the Council of the Indies and the Secretaries of State), Ultramar (the nineteenth-century central agency for the colonies); printouts, such as the inventories of Consulados (Boards of Trade from Seville and Cádiz) or Correos (the Post) and some catalogs of maps and plans; loose file cards (catalogs of Registers); and bound volumes.

In some cases, individual documents are described (e.g., *Patronato* inventory), while in others only information at the series level is covered (e.g., Government Section inventories). The elements of description used also vary considerably—from the title, reference number, and end dates used exclusively in many inventories to the detailed description appearing in catalogs such as the *Catálogo de Consultas del Consejo de Indias* or the *Catálogo de Pasajeros a Indias*.

Some catalogs and inventories include supplemental indices, while others do not. The indices themselves are diverse and do not conform to any established norm. Some contain long headings that virtually summarize the referenced document (for example, supplemental indices for the Contaduría or Contratación sections); other indices are simpler, including only the indispensable data. Some cite the document reference number, others the corresponding page of the principal instrument, and still others (for publications) a consecutive number identifying each entry or individual description.

In short, there is a wide range of finding aids with different objectives, formats, and criteria. A study conducted before the retrospective conversion tasks were begun estimated that all of the finding aids totaled 25,000 pages. Many of them, however, were not available for use by researchers in the Reading Room.

[16] The most significant Archivo guides are those by José Torre Revello, *El Archivo General de Indias de Sevilla. Historia y Clasificación de sus Fondos* [The Archivo General de Indias of Seville: History and Classification of its Holdings] (Buenos Aires: Instituto de Investigaciones Históricas, 1929), and José María de la Peña y Cámara, *Archivo General de Indias. Guía del Visitante* [Visitor's Guide] (Valencia: Dirección General de Archives y Bibliotecas, 1958).

[17] Some of the most recently published inventories include the Consulados or Correos (Boards of Trade or Post) sections.

[18] For example, the *Catálogos de Pasajeros a Indias* [Lists of Passengers to the Indies] (7 vols. published), or the many catalogs of maps and plans by Pedro Torres Lanzas, Julio González, or María Antonia Colomar.

Objectives for the Information and Reference System

- To construct a database containing valid information for locating digitized documents and to evolve a unified, global system of archival information that could handle all descriptive data of the AGI in an integrated manner. All the traditional finding aids would have to fit within the new system, so that the entire process of description and search for information could be integrated and automated. This unified system is to some extent consistent with the aim of a "general inventory" set forth in the Ordenanzas.
- To include in this unified system the information needed for better retrieval of the images of digitized documents. Such data would not be independent but an integral part of the system as a whole. Yet a considerable effort would have to be made to supplement and enhance the description of that documentation with a view to better results and more rapid retrieval.
- To undertake complete retrospective conversion of the large volume of guides, catalogs, and indices that had been developed, despite their errors and lack of standardization.
- To ensure that the new system respects traditional archival principles and the "principle of provenance" in particular. This meant organizing the descriptive information according to the organic-functional structure of the fonds, which calls for a hierarchical and multilevel model of data access.
- In addition, to allow direct access to the information through the use of keywords, as an alternative to access through the hierarchical path.
- To offer vocabulary control options for the future, although the AGI traditionally had no standardization or control for retrieval vocabulary (subject headings, thesaurus, control of authorities).
- To simplify integration with the user-management and image modules by incorporating all necessary elements—such as data access controls and references to new media—into the information and reference system.

Outline of Information and Reference System

Given these objectives, the information and reference module was designed to construct a unified data system that would make it possible to access information through the "principle of provenance path," while also providing new possibilities for direct data access through new technologies.

Data Access by the "Principle of Provenance Path"

The system manages relational model databases to set up a hierarchical and family-tree model of all descriptive information that is fully consistent with the AGI's organic-functional structure. Following this

framework, one can navigate from the "root" or holding institution (the Archivo) through its several branches to its "leaves" (documentary units or pieces).

Holdings are organized by level, which may or may not be fully reflected in the finding aids. The information system should allow management within this structure of existing or described "real" levels. This is not a rigid information structure with preestablished levels. The system allows the inclusion and management of as many levels as may be considered necessary, depending on the requirements and holdings of the archive.

The International Standard Archival Description, ISAD(G), reflects this approach, stipulating in rule 2.3 that the system of organization and access to information should "link each description to its next higher unit of description, if applicable, and identify the level of description." But this standard, which was adopted some time after the AGI computerized system was developed, does not specify a means for its implementation.[19] The AGI information and reference system employs a simple procedure: if each descriptive entry, at whatever level, has an identification code, only the code of the next higher unit in the hierarchy needs to be included in the database table to link both units.

Since the user ordinarily does not know the internal identification code, and it is also difficult to remember, the Reference Number or code of each description unit is used for the purpose. Thus, this operation in the data entry process introduces a new "element of description" that is always necessary: the Reference Number or code of the higher hierarchical unit. During consultation, the user has the option of going to a higher or lower hierarchical level. The descriptive entry corresponding to a section will indicate its component series, or a document will show all the levels above it.

Direct Information Access

AGI Indices

The principle of provenance, which provided the main means for access to information at the AGI, is an indirect method of obtaining materials. Document access is indicated by the path (search subject/producing institution/inventories and catalogs/supplemental indices), rather than by the more direct subject/document route.

To provide points of more direct access, the AGI developed indices to supplement the principal finding aids. The indices represent document content through different "notions" or "concepts" organized in the form of index "headings." They may be direct (index heading/document) or indirect (index heading/corresponding heading in its inventory or catalog/document).

[19] International Council on Archives. ISAD(G): General International Standard Archival Description. Adopted by the Ad Hoc Commission on Descriptive Standards. Stockholm, Sweden, January 21-23, 1993.

In recent years, efforts have been made in the archival field, both nationally and internationally, to create standards for indices production. For example, the International Council on Archives (ICA), through its Ad Hoc Commission on Descriptive Standards, has drawn up an International Standard on Authorized Archival Headings for Organizations, Persons and Families ISAAR (CPF), which was adopted in Paris, November 15-20, 1995.[20] Also, the Bureau of Canadian Archivists, through its Planning Committee on Descriptive Standards, published several related texts—such as the Subject Indexing for Archives, Ottawa, 1992—that contain bibliographies and references to existing standards.[21]

The AGI had no vocabulary control. In the planning phase, there was discussion of whether to use existing indices or to revise them. Revision would require the preparation of strict rules of control, or a thesaurus, or list of subject headings. Project staff agreed that it would be impossible to revise the indices properly within a reasonable time. Within the framework of the digitization project, it was impossible to draw up a complete list of acceptable headings valid for all sections and all levels of the AGI. The differences and incongruities among existing indices were too great. Thus, existing indices were used, with no vocabulary control in the strict sense. In the process of retrospective conversion, they were converted to the actual keywords that provide direct access to the information.

An important area of work for the future could be to establish some kind of controlled vocabulary, employing for the first time lists of all index headings used in the AGI, now available as keywords in electronic format.

Keywords

Keywords are the principal means for direct retrieval of information from the system. They may consist of one or more words, providing they do not exceed 120 characters. They include indices by name, toponym, institution, and subject.

These keywords can be interrelated by means of precoordinated indexing. When a unit of description is indexed, two different keywords may be linked by means of a specific relationship, such as nature, affiliation, title, or activity. For example,

First Keyword	*Relationship*	*Second Keyword*
Pizzaro, Franciso	Born in	Trujillo (Cáceres)
Cortés, Hernán	Title	Marqués del Valle
Colón, Diego	Son of	Colón, Cristóbal
Gálvez, Barnardo de	Activity	Gobernador de Luisiana

[20] ISAAR (CPF): International Standard Archival Authority Record for Corporate Bodies, Persons, and Families. Final ICA approved version. Prepared by the Ad Hoc Commission on Descriptive Standards, Paris, November 15-20, 1995.

[21] Bureau of Canadian Archivists. Planning Committee on Descriptive Standards. Subject Indexing for Archives: The Report of the Subject Indexing Working Group. (Ottawa: Bureau of Canadian Archivisits, 1992).

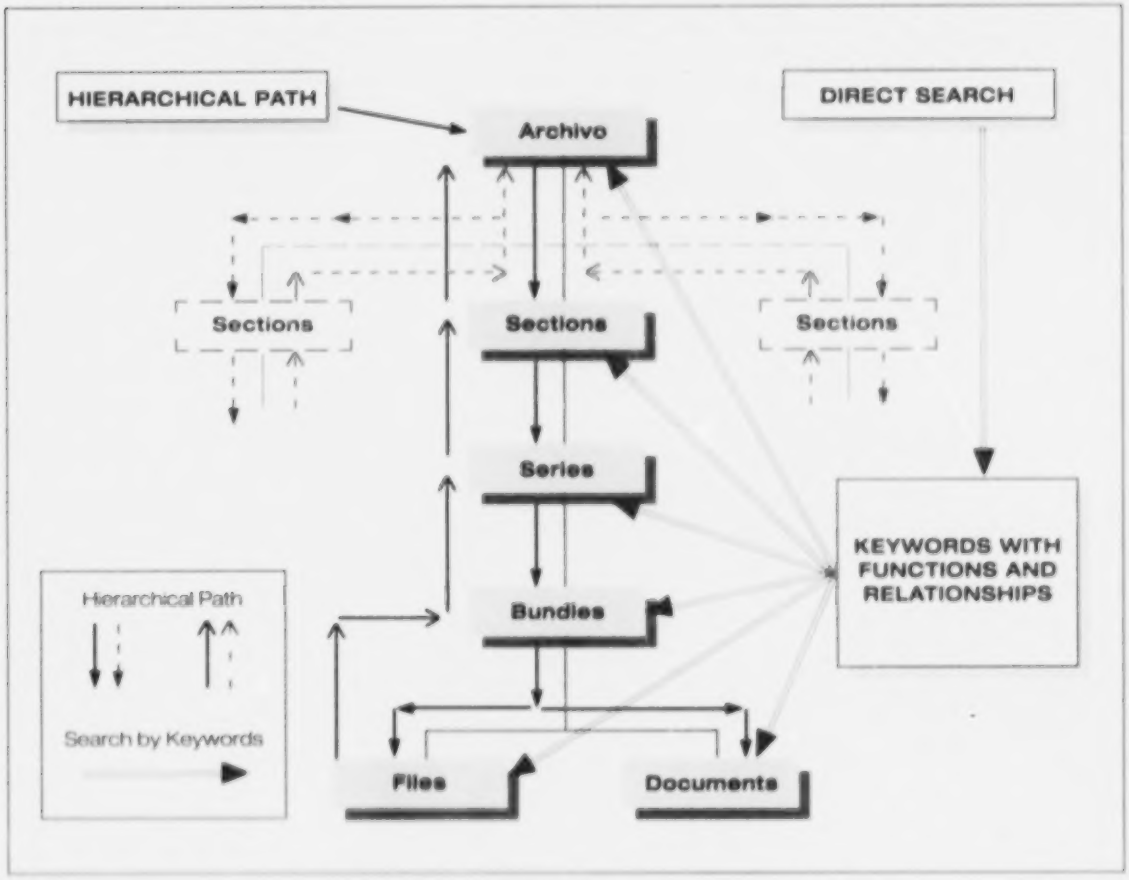

Figure 3. *Means of Access to Information*

They may also be "qualified" by the function they perform within the corresponding document. For example, one person is the "otorgante" (giver) of the last will, the "witness" in a judgment, or the "receiver" of a letter. The name of the person is the keyword, which can be qualified with the function of "otorgante", "witness", "sender", or "receiver."

In the process of consultation, if one or more search words are introduced, the system will show all of the keywords in the database that contain those words, making it possible to locate the descriptive information about a particular document or series of documents. For each descriptive entry, all keywords considered useful may be included, with or without specific relationships or functions.

Although there is no compulsory vocabulary control, the system has the capacity to use a thesaurus as an element of support and consultation. It is also supplied with a specific profile of "responsible" vocabulary control unit, which can carry out revision functions such as the elimination and creation of keywords, elimination and creation of specific relationships, and transfer of references between keywords and relationships.

BEST COPY AVAILABLE

Elements of Description

Once a decision had been made on the structure for the unified system of information and the major means of access to documents, it was necessary to select the various elements of description (title, content summary, volume, date spans, etc.) to ensure adequate gathering of all information concerning AGI documents. This was some years before the EAD (Encoded Archival Description) was developed. Certain existing possibilities were studied, including the MARC AMC format, but it was decided to choose something simpler, adapted to the AGI's specific needs. At the time there was no internationally recognized standard for archival description, although Canada, Britain, and the United States had developed some national standards.[22] In 1992, International Standard Archival Description, ISAD(G), was proposed to the professional community at the International Congress on Archives, held in Montreal. It was subsequently adopted by the Ad Hoc Commission at its Stockholm meeting, January 21-23, 1993.[23]

Consequently, a data structure was developed that consists of 30 elements of description divided into three areas (basic information, descriptive information, and elements of retrieval). The main data are summarized below. A numeric label identifies each of the several elements.

Basic Information Area

All of the elements contained in this area are required, although during data entry most can be included as values by default. These elements include the reference number, dates, title, and reference number of the preceding element in the hierarchy. Almost all of this corresponds to the Identity Statement Area of the ISAD(G) standard, although additional elements needed for service are also included.

Descriptive Information Area

This covers a series of elements of description that are optional, although some are used in almost every case, while others are used for specific documentary types. These descriptions are usually in a free-text format. Examples of the elements are the content summary, old and new reference numbers, site of issue, and internal and external characteristics.

[22] Steven Hensen, *Archives, Personal Papers and Manuscripts: A Cataloging Manual for Archival Repositories, Historical Societies and Manuscripts Libraries* (2nd ed.) (Chicago: Society of American Archivists, 1989); Michael Cook and Margaret Procter, *Manual of Archival Description (MAD)* (London: Society of Archivists, 1989); Bureau of Canadian Archivists, *Rules of Archival Description* (Ottawa: Bureau of Canadian Archivists, 1990).

[23] International Council on Archives, *General International Standard Archival Description*, ISAD(G), adopted by the Ad Hoc Commission on Descriptive Standards at its Stockholm meeting, January 21-23, 1993.

Elements of Location, Retrieval, or Access

These are the elements facilitating access to the data, including the keywords and their several supplements, such as specific relationships and functions. Also included are such elements of retrieval as the old card numbers or those in other media, additional dates to focus searches, and so forth.

Initial Content: Retrospective Conversion of Finding Aids

At the beginning of the project, two distinct possibilities were studied regarding the initial content of the information and reference system:

- To incorporate only the descriptive information necessary for retrieval of those documents included in the digital image storage system; or
- To incorporate all information contained in the AGI's finding aids; that is, inventories, catalogs, and indices, whether printed, typed, or in manuscript.

In the first case, a database structure could have been designed that was appropriate for the documents to be digitized. In addition, efforts could have been made to produce a more precise description of only a subset of AGI documents, determining the fields necessary for each document type.

The second option, which was the one ultimately chosen, was a new venture entailing considerable risk: there was no previous experience to serve as reference, since no archive had yet undertaken a similar operation. The problems—the quality of existing descriptive information, different levels of description, lack of uniform criteria and standards over time, and different formats—posed several difficulties.

A complete list was made of all the finding aids to be converted (estimated then at 25,000 pages). The content and format of each was analyzed, and priorities were established for the subsequent work, to be undertaken in phases:

- Finding aids essential in forming the complete hierarchical structure of the database were assigned priority.
- Within that group, special emphasis was placed on finding aids favoring preparation of the material to be digitized (for example, that included in the Patronato Section, all of which would be computerized).
- Priority was then attached to all catalogs and indices describing certain resources or series in detail and in depth.

IBM's Personal Editor was adapted for the work to provide plain data entry screens consistent with the needs of each of the finding

aids to be recorded. The data were stored in simple flat ASCII files, with identification codes or labels for each of the fields employed and with control words adding information needed at the time of data entry. Two work groups were organized:

- Finding aids published or properly prepared and structured in typescript were keyed in by personnel with no specialized archival background but with data entry experience.
- The finding aids most difficult to interpret (usually manuscript) and requiring more careful review were recorded directly in the AGI by a group with more extensive archival and historical training, overseen directly by the regular staff.

The results of the process can be outlined in summary as follows:

- The idea of the unified information system took shape as a real operating system following an extended, ongoing effort.
- Continuity of the project with the same guidelines over a long period allowed successful completion of the original aims.
- The goal was extremely ambitious, but it was always focused on doing whatever could be done at the time.
- The decision to go public with a system filled with as much content as possible, as opposed to a database structure without content, led to acceptance by researchers.
- Based on this introductory information, the initial errors and inconsistencies could provide incentive to continue the effort, with the goal of gradually supplementing and improving all of the AGI's descriptive information.

The Information and Reference System Today

This retrospective conversion operation has been complemented by a special effort to improve and expand descriptions, especially for the digitized documents. The AGI's information and reference system currently includes the following breakdown of entries for each of its "units of description":

- 37 entries at the section level
- 440 at the series level
- 187 at the subseries level
- 45,398 at the bundle level
- 113,936 at the file level
- 193,849 at the document level
- 61,519 entries for passenger lists to the Americas

These hierarchical entries have been supplemented by about 400,000 keywords to facilitate direct access to specific documents.

Today, we can make some observations on the information and reference system with a view more to the future than the past, draw-

ing on five years of service to users and recent developments in both technology and the development of standards.

Direct and Indirect Access to Archival Information

For many years, the AGI Reading Room has been an excellent field of observation for analyzing the means of intellectual access to archival information because researchers could use both the traditional means of access, by way of the principle of provenance, and the direct means, through indexes.

Although it is risky to reach conclusions because there have been no statistical studies in this area, certain statements can be made based on daily observation. For example, it can be said that a researcher, especially one with little experience in using archives, makes primary use of the direct access system (keywords), usually ignoring the indirect (principle of provenance, hierarchical path), which is always slower and more tedious. Such researchers are often impressed by the easy access to initial data, which makes them believe they have retrieved all available information, even though only a small percentage of the total AGI has been concisely indexed.

Richard H. Lytle undertook an experiment in this area in 1978 and published the results in 1980.[24] One of his conclusions—that the indexing method offered a greater variation in results (better and worse)—is apparent in this connection. If indexing is complete, it is undoubtedly a rapid, useful method for the user. If not, search results will be poor. Clearly, this may have "silenced" a large volume of documents that have barely been covered by a general description and are not indexed.

But in addition, part of the document context information may be lost through exclusive use of indexes. It is well known that an archival document is not an individual, independent piece but acquires its meaning from the environment in which it has been produced. Thus, the traditional method of searching continues to be not only valid but essential. The complementary use of both methods is, today, still the best means of access to information.

The staff responsible for information must therefore guide the less expert researcher. Nevertheless, one veteran researcher said that he had located in a few days more information than he had found in weeks of work before the new system was installed.

Information Retrieval Tools

One achievement of the project has been to develop a unified system of descriptive information for the AGI. But the tools used to build the system now urgently require updating. Use of a system of relational database management with SQL was a significant advance when the relational model was still used basically for totally structured information and management situations. But today, the system suffers from excessive rigidity, and not all of the advantages offered by un-

[24] Richard H. Lytle, "Intellectual Access to Archives," *American Archivist* 43 (1980): 64-76 and 191-208.

structured information search systems—such as free text, full text, and truncations—have been used. In this regard, the information system has lagged far behind and urgently needs updating.

Standardization

Standardization of all aspects of the information and reference system is another crucial issue for the future.

- Standardized description is the first aspect to be considered. In the near future, it therefore seems advisable to incorporate at least the ISAD(G) standards, although they are still a general and imperfect effort. They are also far removed from other developments in the field of documentation, such as the ISBD or AACR2. It will be necessary to monitor such recent developments as EAD, implemented in the United States through the Berkeley Finding Aid Project and supported in the SGML standard.
- It is also necessary to standardize the language of indexing. This will be a long-term project for the AGI, since no list of acceptable subject headings or indices has been developed. The advantage is that the AGI can do this *a posteriori*, using not only international standards in the field but all keywords generated over the course of two centuries, which are currently available in a unified, automated listing.
- Standardization of other aspects, such as the system for querying databases (SQL, ODBC) or Internet access, must also be addressed.

Revision and Updating of Information System Content

Archival work consists, ultimately, of furthering the organization, description, and understanding of documents. Similarly, the AGI's information system will require continuous revision, updating, and improvement regarding both description and access. In the process, it will be possible gradually to eliminate errors originating in the retrospective conversion.

DIGITAL IMAGE STORAGE SYSTEM

The ability to view and consult on-screen many original documents through their digital images stored on optical disk is the most innovative and well-known aspect of the project.[25]

As noted earlier, the AGI has two priorities: the sound conservation of original documents, and the optimal access to those documents by researchers. In 1992, the Reading Room accommodated some 50 researchers each day. In 1990, 37,303 consultations of original documents were recorded, and in 1991, 37,172.

Since this is a relatively small archive (slightly more than 43,000 document bundles), such heavy use by researchers and photocopiers entails a serious risk of deterioration of the original documents. Between the second half of 1989 and the end of 1992, almost 200 bundles in the AGI were consulted more than 50 times in the Reading Room.

How could the dilemma of conservation and document access be resolved? The best alternative was offered by the systems of document reproduction and the use of copies, rather than originals, for consultation. This leads to an important point that affected subsequent technical decisions: *the aim was not to replace the original documents with digital copies but to preserve the original better by avoiding its continual handling during consultation.* Thus, the technical specifications for digitization were set to ensure adequate quality for consultation—not to replace the original for all purposes. The level of quality had to satisfy most researchers so that originals would need to be consulted only in exceptional cases.

Digitization Process

We can analyze the system of digital image storage by dividing it into two basic processes: the process of digitization and storage on optical medium, and the process of consultation. Digitization in turn can be studied in two further steps: the preparation of documents or prior archival work, and the digitization itself.

Selection of Documents to be Digitized

Since it was not possible to digitize all of the AGI documents in a reasonable amount of time, the first question asked was how to make proper selection of the documents to be scanned. The following selection criteria were used:

- Only complete series would be digitized—never selected individual documents. This was the first and mandatory criterion.

[25] Pedro González, "¿Salas de Lectura sin Papel?" [Paperless Reading Rooms?], in *Proceedings of the 11th International Congress on Archives* (Munich, New York, London, Paris: K.G. Sauer, 1989), 229-33.

- Documents found to be of greatest use for consultation would be selected. A statistical analysis would locate the documentary series that researchers had most often used.
- The documents selected would cover all territories relating to Spanish colonization in the New World. This would draw on the Archivo's strength as the basic archive for history of the Americas.
- As a practical criterion, the status of document description was also considered, with a view to the work of preliminary preparation.

The final selection yielded a list of documentary series that would satisfy more than 30 percent of users' requests. The list represented about 10 percent of all AGI documents. It was revised by adding some documents held by two other historical archives relating to Spanish history in the New World, the Archivo Histórico Nacional (AHN) and the Archivo General de Simanacas (AGS). These documents, however, were not selected by the same criteria as those for AGI.

The purpose of supplementing the Archivo's holdings was to meet fully the goal established by King Carlos III when the AGI was founded: that all documents "referring to the Indies" should be deposited in the AGI. A further goal was to incorporate other national archives into the process of technological modernization. More detail about this work appears in Appendix 3.

Document Preparation

The documents selected must be properly prepared for forwarding to the digitization room. This entails the traditional work of organizing the documents, bundle by bundle, drafting new descriptions for them or revising present ones, placing them properly in folders, writing the reference number, etc. These steps are fundamental in ensuring a subsequent expeditious and successful retrieval of digital images.

The final step in preparation is entry into the information system of the descriptive data that will eventually enable document access.[26] A "digitization guide" is also prepared, a form that includes the minimum information needed to guide the scanner operator in his work and to carry out the subsequent liaison between the textual database forming the information and reference system and the digital images stored on optical disks.

[26] "Efficient retrieval of scanned document images and graphic data depends on the accurate, up-to-date index data base. Indexing a digital image involves linking descriptive image information with header file information... And accuracy is critical because an erroneous index term may result in non-retrieval of the related image." National Archives and Records Administration (NARA). *Digital Imaging and Optical Digital Data Disk Storage Systems: Long-Term Access Strategies for Federal Agencies.* Technical Information Paper no. 12 (NARA, 1994).

Digitization

Once the documents have been prepared, they are digitized. From 1989 to 1994, digitization stations were supplied with an IBM AT computer, a Rank Xerox 7650 flatbed scanner, and an optical disk unit—first an IBM 3363 and later a Panasonic.[27] The scanner could digitize up to size A3 to 400 dots per inch (dpi) with 256 grayscale levels, although the AGI uses only 100 dpi and preserves only the most significant 16 grayscale levels.

The same routine is followed for each page: scanning, viewing, compression, and direct recording on optical disk. The entire process takes about one minute per page. The documents—because of their intrinsic value, age, and state of conservation—require careful handling during digitization, so no automatic scanner feeding is allowed. From 1990 to 1992, the work was carried out by 15 digitization stations working double shifts with a variety of equipment to prevent delays in case of breakdown. The number of staff decreased between 1992 and 1994, when the agreement between the three partners ended. In 1995 and 1996, work continued with three full-time and six part-time employees.

At the end of 1996, the AGI got two new digitization stations, each equipped with a Kodak DCS 420 digital camera and Hewlett-Packard CD-Writer 4020i disk recording units. This made it easier and faster to complete the work (only seconds per page), with less risk to the documents. Now, AGI requires only one-third the staff to achieve the same rate of productivity as with the old equipment.

The most important aspects of AGI's digitization work are image quality, storage support, formats and image compression, and quality control.

Image Quality

The quality of a digital image is determined basically by two parameters: resolution and grayscale. How can the AGI obtain a digital image offering quality guarantees for its stated goals? In recent years, a series of projects have tried to set standards guiding the selection of such parameters, and various criteria have been suggested. Don Willis distinguishes between "archival resolution," defined as "the resolution required to ensure a faithful replica of the original document regardless of cost" (with archival resolution at 600 dpi and 8 bits of grayscale), and "optimal archival resolution," defined as "the highest resolution economically supported by technology at a given time" (assuming a balance between cost and quality), and "adequate access" resolution, disregarding conservation but focusing on information needs (estimated at 300 dpi in black and white).[28]

---

[27] The Panasonic optical disk was marketed by the British firm Plassmon, under the trademark Reflection Systems RF-5010C.

[28] Don Willis. *A Hybrid Systems Approach to Preservation of Printed Materials.* (Washington, DC: Commission on Preservation and Access, 1992), 11.

Obviously, many types of documentation can be digitized, each with its own characteristics and requirements.[29] Anne R. Kenney and Stephen Chapman divide the documents by category: text or line, halftone, continuous tone, or mixed documents. Each type requires different parameters in principle. While some maintain that "archival resolution" is 600 dpi, NARA proposes 300 dpi, after having used 200 dpi in its ODISS (Optical Digital Image Storage System) project.[30]

Different projects have different goals that determine digitization parameters. Seeking complete "replacement" of the original by the digital image differs from digitizing solely for on-screen consultation or on-line network access to limit handling of originals.

Document characteristics also determine the selection of parameters. There is a difference between digitizing modern typed documents that are well preserved and documents of the sixteenth or seventeenth centuries with bleedthrough from ferrogallic inks or with inks faded by exposure to humidity. A suitable balance among costs, budget resources, and the stated objectives must also be considered. Ignoring these aspects while pursuing optimum quality could make it impossible to maintain projects because of the implications for storage and even general systems configuration. For example, optimal quality is likely to require more powerful processors, higher-quality monitors, and enhanced capacity networks.

In terms of the AGI project, the aim is to obtain reference quality that will allow users to consult the document on-screen or in hard copy, rather than in its original form. The documents to be digitized are manuscripts from the fifteenth to nineteenth centuries, in various states of conservation. They may show water spots, stains, faded ink, or bleedthrough. They can be classified as "continuous tone" according to the classification drawn up by Anne Kenney and Stephen Chapman.[31]

The fundamental criterion is the search for a proper balance between image quality on the one hand and storage and processing-capacity needs on the other. Fortunately, the costs of storage and processing capacity have dropped steadily, making it less expensive to store and process good quality images. Technological advances will also lead to the development of reasonably priced equipment that can handle vast quantities of information.

Several tests were conducted to ensure quality consistent with the purpose and type of documents,[32] with an eye to minimizing

[29] Anne R. Kenney and Stephen Chapman. *Tutorial: Digital Resolution Requirements for Replacing Text-Based Material: Methods for Benchmarking Image Quality* (Washington, DC: Commission on Preservation and Access, 1995).

[30] National Archives and Records Administration. *Digital-Imaging and Optical Digital Data*.

[31] Kenney and Chapman, *Tutorial*, 2.

[32] See Julián Bescós and Juan Navarro, "La Digitalización como Medio para la Preservación y Acceso a la Información en Archivos y Bibliotecas" [Digitization as a Means of Preserving and Accessing Information in Archives and Libraries], *Educación y Bibliotecas* 80 (1987): 28-41.

storage requirements. Based on these tests, it was decided to digitze at 100 dpi and 16 grayscale levels (4 bits per pixel). These parameters are quite different from those used in other digitization projects but are well suited to the aims of the Seville project—to digitize for consultation, rather than for replacement of the original. The parameters are also consistent with the minimum values recommended in 1995 by the Technology Subcommittee of the UNESCO Memory of the World Programme.[33]

Based on these parameters, project staff decided not to improve images during digitization, but to defer that step to the stage of viewing. The viewer decides on the use of image enhancement algorithms at the time of consultation. This procedure also simplifies and shortens the work of digitizing: the operator does not need to make any special decisions regarding image quality, which is assumed to be adequate if the specified parameters are maintained.

The effect of digitization with grayscale on resolution requires further study, as noted by Kenney and Chapman.[34] Their comments are made with respect to tests that the Cornell Department of Preservation and Conservation conducted on brittle books. The impact of digitization with grayscale on the cost of storage, time, equipment, and processors is declining steadily owing to the rapid growth of processing capacity.

Color Images

The project also aimed to digitize AGI's maps and plans collection, which consists of about 7,000 pieces that have been extracted from their original, provenance-based files to facilitate their conservation. These materials differ from other normal documents in that they use color and are oversized. Because of this, AGI had to use computers with greater capacity and processing speed, and new solutions for storage had to be found. It was also impossible to use the same scanner as for other materials.

It was decided to first make a color microfilm copy, which would subsequently be digitized. Microfilming with Cibachrome, at 200 line pairs, would yield a good-quality color copy. The film could then be digitized with a Nikon LS3500 slide scanner, able to capture 4,096 x 6,144 pixels for 35 mm frames. The final resolution in digital form was similar to the grayscale documents (100 dpi), with 256 colors. For quicker access, a black-and-white copy could be saved.

The color microfilming was completed, but the scanning operation was interrupted and only one hundred digital color images were obtained. There were three reasons for this. First, although the quality of the images appeared basically good, sometimes the resolution was too low for sharp text legibility. Second, it took too long to display the large color document images on the screen with the 486 pro-

[33] Stephen Foster, Roslyn Russel, Jan Lyall, and Duncan Marshall. *Memory of the World: General Guidelines to Safeguard Documentary Heritage*. Report CII-95/WS-11 (Paris: UNESCO, 1995), 52-66.

[34] Kenney and Chapman, *Tutorial*, 10.

cessors. Third, this work assumed lower priority after the AGI could offer consultation of maps and plans through color microfilm in the Reading Room.

Storage Medium

Most of the AGI documents are folio size, measuring slightly over A4. An A4-size image, digitized at 100 dpi and 4 bits/pixel uncompressed, can occupy about half a megabyte. While still considered sizeable in 1997, the storage requirement was much more significant ten years earlier.

Besides needing image compression algorithms, there were few choices for media carriers. The most useful option was the WORM (write once/read many) optical disk because of its capacity, ease of recording and of subsequent use for consultation, and predictable longevity. The fundamental problem with WORM disks was the absence of standards and the variety of trademarks and formats, which increased the risk of obsolescence.

A 200 MB capacity IBM 3363 optical disk was used first. This disk model, on which 1,729 bundles were digitized, was soon replaced by a Panasonic (Reflection Systems RF-5010C) disk with 940 MB capacity. The Panasonic could usually store all the images of a bundle on a single disk, after compression. It was used to record 3,732 bundles of documents. Each new RF-5010C disk cost 15,000 pesetas wholesale (about US $100). Thus, with 5,511 bundles digitized by the end of May 1997, the cost of storage media was very high. Assuming that each disk held the images of one bundle, and each bundle averaged 1,956 pages, the per-page storage cost could be estimated at 7.67 pesetas (approximately five cents).

The process of converting all disks to the new media carrier began with the changeover from the IBM to Panasonic disks. (The conversion was not completed as of 1997 and the consulting system was still using both types of disks at that time.)

In recent years, the use of CD-ROM has become more common. CD-R recording equipment has been developed, and the cost of blank CDs has dropped. Consequently, it was decided to migrate to this new format. At the end of 1996, the AGI acquired new recording equipment with CD-R units and installed six units for converting WORM disks to CD-R (three to convert IBM disks and three for Panasonic disks). Migration has been under way since early 1997 and should take about two years.

For several years, the lack of image backup copies has been a key problem. The AGI postponed a decision on the backup system in anticipation of better options, such as magnetic or optical tape. Meanwhile, the AGI had only a single copy of the images, risking the loss or deterioration of a disk, which would require redigitization of the original. A backup copy program was finally established in 1995. Five units were installed for copying Panasonic disks onto DAT DSS I and II magnetic tape (2 and 4 Gb tapes). Between April 1995 and the end of 1996, 3,205 disks were copied, leaving only 525 IBM disks uncopied onto a new media carrier. These are now being copied di-

rectly to CD-R in the new migration process that began early in 1997. Now, when documents are digitized, two CD-R copies of the digitized images are made: one for use and another for backup. The cost of a new CD was $6.50 at the beginning of 1997, but is now much cheaper.

By the end of May 1997, the status of migration was as follows:

IBM disks:

- Digitized: 1,729 bundles
- Copied on Panasonic: 1,186 bundles
- Copied on CD-R: 110 bundles
- Uncopied: 433 bundles

Panasonic RF-5010C disks:

- Digitized: 3,732 bundles
- Duplicated on Panasonic: 571 bundles
- Copied on magnetic tape: 3,205 bundles
- Copied on CD-R: 141 bundles

CD-R disks:

- Digitized: 50 bundles (two copies)
- Copied from IBM: 110 bundles
- Copied from Panasonic: 141 bundles

Magnetic tape:

- Copied from Panasonic: 3,205 bundles

Formats and Image Compression

The heavy storage requirements for digital document images call for the use of compression formats. The AGI wanted to scan for grayscale, rather than for black-and-white or binary images. But standard algorithms for color or grayscale images, such as JPEG and GIF, were not available in 1988. Consequently, the AGI digitization project developed its own compression algorithm using a DPCM (Differential Pulse Code Modulation) model with statistical compression. This model allows for an approximate reduction of 2 to 1 without loss of quality.

The increased use of JPEG made it advisable for the AGI to eventually adopt this compression algorithm. Current migration to the new CD-R support is also changing the compression format. Decompression of the AGI format and subsequent compression in JPEG are carried out during the recording process. Since JPEG allows adjustment of the compression parameters to determine the quality sought in viewing, the new compression is designed to permit no more than a 15 percent loss. The compression factor is similar to the AGI for-

mat, with the images occupying the same amount of space in both AGI and JPEG formats.

Quality Control

After digitization, quality control can be done in two ways. The first is automatic, by comparing the digitization guide with the resulting optical disk. This process can detect mistakes, such as omissions and reference number errors, that can be resolved before the WORM disk is dispatched for consultation service or before transfer from hard disk to the CD.

The second means of quality control is manual inspection. This can be done by accessing document images at selected intervals (such as every page, or the first pages of each document or block, one page of every five, or one page of every day's batch of digitized images). In practice, this type of quality control creates a bottleneck since it must be done by specialized personnel and is time-consuming. It was done for only a brief period at the AGI. The users themselves—outside researchers and staffers—have been the ones to detect possible errors, such as omissions or repetitions.

Once digitization and quality control with the first method have been completed, the resulting disk may be used for consultation. It may also be duplicated and a backup copy made.

Image Consultation

Disk Service

In deciding how to place such a large volume of images in service, several options were considered.

- **Decentralized service** would allow users to retrieve the disks themselves, but it has several drawbacks. These include security risks and the fact that readers need two types of optical disks for each user station. There would also be the problem of many researchers simultaneously searching for their disks. This option would be more attractive if few disks were involved.

- **A centralized strategy** called for various possibilities to be studied:

  - **Use of a jukebox**. This is the usual solution for centralized service. But no single unit could handle so many optical disks, and the use of a battery of jukeboxes greatly increased the cost. In addition, a large space was needed to house them.[35]

[35] Jukeboxes on the market now can handle several hundred CDs, which means they could be practical, especially if the new DVD standard is adopted. This would yield much more storage capacity than the current CD and a sizable reduction in the number of disks needed.

◊ **Use of a robot**. There were no robots available on the market for optical disks, only for magnetic tape cartridges. One of these could have been adapted, but the cost was very high and the risk of obsolescence significant.

◊ **Centralized service with human intervention.** A user's request for a disk would be handled by a human operator, located in the Optical Disk Room, who would retrieve the disk and place it in the reading unit.

The use of a set of high-capacity magnetic disks was not studied because they were considered impossible to obtain at the time. This option now begins to offer possibilities, and its use will probably be justified for urgent requests (over the Internet, for instance).

The AGI chose the option of centralized service with human intervention. It was the least expensive and allows for the introduction of other options in the future. Optical disk servers were installed in the AGI Optical Disk Room, connecting more than a dozen optical disk readers for the two existing formats. All available optical disks are installed and organized in shelves beside the servers and the optical disk readers. An operator at that site handles the requests for images as shown on a monitor displaying disk request messages.

There are now different servers for IBM and for Panasonic disks. It will eventually be necessary to install a server, with reading units, for CDs.

When a user at a workstation requests an image, the system generates a message shown on the monitor. The operator receives the message, selects the requested disk from its shelf, and places it in an available disk unit. The document images are sent through the local area network (16 megabits per sec.) to the user's workstation. When all the document images requested have been delivered to the workstation, the optical disk may be withdrawn by the disk unit operator, who is then ready for the next request. The entire process can be carried out relatively easily and efficiently by a single person handling requests through the monitor and a few disk-reading units.

This means that part of the consulting process is not automated and that human errors can occur. The time required for sending the images is brief: in a minute to a minute and a half, the researcher will receive the first image on a monitor and the series of pages constituting the document or file requested will begin to be stored in the workstation. Within a short time, the researcher will have the entire document. From that point on, the researcher can work locally for as long as desired.

Image Consultation

When the researcher receives the requested document images on the monitor, he or she can begin to consult them, using a variety of tools for image treatment and enhancement.[36]

In the original version of the system, the workstation consisted of a PS/2 computer with a 486 processor and OS/2 operating system, and with Dialog Manager and Presentation Manager for user interface. These controlled two monitors: one conventional VGA (IBM 8513) for text and image management, and another high-resolution unit for image display (IBM 8508 for grayscale images or IBM 6091 for color). This initial interface with two screens has been modified; the subsequent versions carry out all functions in a microcomputer with a Pentium processor and a single monitor. The same mix of facilities is retained for document management, browsing, expansion and rotation, printing, and the use of algorithms for historical document treatment (such as elimination of stains and ink bleedthrough, enhancement of faded inks, and improvement of contrast).[37]

[36] For research conducted to develop image enhancement tools for better readability of documents, see:

Julián Bescós Ramón, "Image Processing Algorithms for Readability Enhancement of Old Manuscripts," in *Electronic Imaging 89* (Pasadena, CA, 1989), 1:392-97.

Julián Bescós Ramón, Francisco Jaque, and Luis Montoto, "Reflectance and Optical Contrast of Old Manuscripts: Wavelength Dependence," *Scanning Imaging* Vol. 1,028 (Society of Photo-optical Instrumentation Engineers [SPIE], 1989): 258-62.

Julián Bescós Ramón, Juan Pedro Secilla, and Juan Navarro, "Filtering and Compression of Old Manuscripts by Adaptative Processing Techniques," *Proceedings of the Society for Information Display International Symposium 1990* (Las Vegas: Society for Information Display, 1990): 384-87.

Julián Bescós Ramón, Juan Navarro, and Carlos Ramón, "Mejora de Legibilidad de Documentos Antiguos Mediante Tratamiento Digital de Imágenes" [Enhancing Readability of Old Documents through Digital Image Treatment], *IV Simposium Nacional de Reconocimiento de Formas y Análisis de Imágenes* [Proceedings of Fourth National Symposium on Form Recognition and Image Analysis] (Granada: Sociedad Española de Reconocimiento de Formas y Análisis de Imágenes, 1990): 51-58.

[37] In their previously cited report, Hans Rütimann and M. Stuart Lynn note: "The speed and ease of use of these tools are impressive. There is something almost magical in seeing a badly stained section of a 300-year-old manuscript cleaned up before one's eyes and become legible again." Rütimann and Lynn, 11.

BEST COPY AVAILABLE

TECHNICAL ASPECTS

The AGI has had to address the major problem of equipment and system obsolescence. Since the system was devised in the second half of the 1980s, it has become essential to start updating and enhancing hardware and software, and migrating data. The aim of this update has been to incorporate more open formulas, free from "proprietary" influences and adapted to standards subsequently developed or consolidated. The other participants in the original agreement have worked with the AGI in undertaking these operations.[38]

Initial Computer Equipment

The following hardware and software was initially installed in the Archivo General de Indias:

- A 16-megabit-per-second Token-ring local area network with APPC communication protocol for communications support
- An IBM AS-400 minicomputer with SQL-400 as database and user-management system server
- PS/2 microcomputers with OS/2 operating system as optical disk servers, controlling Panasonic and IBM disk units
- PS/2 microcomputers with OS/2 operating system and Dialog Manager and Presentation Manager for user interface as workstations. They control two monitors, one standard VGA (IBM 8513) for text and one high-resolution for image display (IBM 8508 for grayscale or IBM 6091 for color images).
- Digitization stations with IBM AT microcomputers, DOS operating system, Rank Xerox 7650 flatbed scanners, and IBM or Panasonic optical disk units
- IBM 4029 laser printers for printing images
- Stations for digitization of maps and plans consisting of PS/2 model 80s with 16 MB memory, Panasonic (Reflection Systems RF-5010C) optical disk units, and Nikon LS3500 slide scanners

[38] In 1993, during the extension of the initial agreement for the AGI project, IBM España and Informática El Corte Inglés (the principal Spanish company in the field of computers and connected with the Ramón Areces Foundation) decided to establish a new base for the future of the system (continuity, update, enhancement, and diffusion to other archives). Looking to this goal, they signed an agreement to form an association of economic interests officially known as Archivos y Bibliotecas, A.I.E. To date, they have continued the technological development of the system and installed it in about 30 Archives Centers under the trademark ArchiGES (user management system) and ArchiDOC (information and reference and image systems). At the end of 1997, IBM, following a general company strategy toward small business, withdrew. Informática El Corte Inglés continues to maintain and enhance the system.

Current Equipment

The system was designed within a "proprietary" framework with some commercial dependence. This has led to revisions of the system to create more open and interchangeable versions. Accordingly, the initial equipment has been changed over time as the system has been updated. The basic changes in equipment have been as follows:

- Digitization stations with Rank Xerox flatbed scanners have been replaced by others consisting of microcomputers with Pentium processors, Kodak DCS 420 digital cameras, and Hewlett-Packard CD-Writer 4020i disk recording units. Digitization continues at 100 dpi and 16 grayscale with JPEG compression algorithm. The new hardware allows easier and more rapid digitization. This boosts productivity and reduces digitization costs. Moreover, use of the camera instead of the scanner is safer for the original, since there is much less risk of paper deterioration.
- The AS-400 (database server) and PS/2 equipment are still used but can be replaced today by any PC with a standard Pentium processor. Much of the system now has Windows NT available through ArchiGES and ArchiDOC; the rest of the system is expected to run on Windows NT shortly.
- The optical disk and printing servers have been replaced by new Pentium processor microcomputers. Similarly, the workstations—originally two-screen PS/2 models—can be replaced today by conventional PCs with Pentium processors and standard monitors.
- Six WORM to CD-R conversion stations have been added: three converting IBM and three Panasonic disks, with original disk-reading units and Philips CDD 2000 CD-R recording units.
- Five DAT DDS II tape units have been added.

The Token-ring local area network is still in use, as are the microcomputers with 486 and Pentium processors. The AGI has 40 workstations connected to the network.

Other Current Equipment Possibilities

In standardizing and updating the system, it is possible to use several hardware and software options through ArchiGESand ArchiDOC.

- Migration of the entire system has begun for use in operating environments other than OS/2. Much of the system is already available in Windows NT, while DB2 for Unix can manage the databases. Full migration of the Information and Reference and Image systems (ArchiDOC) to Windows NT and Windows 95 platforms will be finished by November 1998. For the User Management System (ArchiGES), full migration will be done in the spring of 1999.

- The system can operate in a standalone workstation or in local area network (Ethernet or Token-ring), with NETBEUI or TCP/IP protocols.
- The system is supported by the DB2 relational model database and is expected to progress to ODBC standard interface for relational databases.
- Several flatbed scanner models can be used for digitization (Ricoh, Bell & Howell, Hewlett-Packard, IBM, and Epson), as can the Kodak digital camera (DCS 200, DCS 420, DCS 460), through SCSI interface. Plans call for the incorporation of new digitization devices.
- Standard compression algorithms have been adopted: JPEG for grayscale images (with selection of compression ratio and loss percentage) and CCITT Group IV for black-and-white.
- The ISAD(G) archival description standard is used in the information system, with free-text consultation to be incorporated in all fields of the standard.
- The development of interfaces for Web clients and servers for the Internet is under way.

NEW PROSPECTS: LONG-DISTANCE ACCESS

During the years when the Archivo General de Indias' computerized system was devised and developed, no priority was assigned to long-distance access. Instead, the emphasis was on development of a system for integral management of the AGI. But some prospects for use of long-distance access solutions were analyzed from the start, even though the explosive growth of Internet use was still far in the future.[39]

The initial designs called for an AGI opening abroad through the Data Processing Center of the Ministry of Culture—specifically, through the Cultural Information Points (PIC), the first publicly accessible cultural network. The network was accessed through the Packet Switching Data Network (PSDN), which has X.25 lines, and the basic telephone network. The PIC network disappeared with the arrival of the Internet, and the information originally mounted on PIC was moved to the Web site of the Ministry of Culture (www.mcu.es). Today, this site includes the *Censo Guía de Archivos*, which offers a database with general information on more than 30,000 archival centers, and the *Bibliografía de Archivos*, where one can consult the professional bibliography for archives.

In 1992, a six-month experiment in long-distance transmission was conducted by installing at the Huntington Library in Pasadena, California, a workstation linked to the AGI that provided access to its descriptive database. This workstation was supplemented by an optical disk unit located at Huntington headquarters that provided access to the digital image of certain selected documents. Thus, the descriptive information could be obtained across the Atlantic, while access to the digital image remained local.

Other possibilities for remote connections were studied, such as through the Spanish academic network (Red Iris). Another experiment was conducted to provide access through the Integrated Services Digital Network; the necessary interfaces were developed, and isolated experiments were done. For example, at the III Jornadas de Administración Pública y Nuevas Tecnologías, held in Palma de Mallorca in May 1995, the AGI documents could be consulted directly following installation of a workstation at workshop headquarters that was connected to the AGI network. But these experiments were not followed up.

The massive emergence of the Internet prompted new thinking about the possibilities for long-distance access, although the communication networks' capacity did not then seem well suited for sending good quality images of the digitized documents. A document consisting of 25 pages (the usual average at the Archivo), for exam-

[39] Pedro González, "Databases and Long Distance Communication. A Spanish Experience." Proceedings of the Third European Conference on Archives. Vienna, 1993. Published in *Mitteilungen des Österreichischen Staatsarchivs* (Sonderband 2, 1996), 319-72.

ple, could require 8 MB. Today, such technological problems are being solved rapidly, making it possible for the AGI to join the "Global Village," offering its holdings through the Internet. But some problems must first be resolved, and certain important decisions made.

Technical Problems

The most important technical problems to be considered are the following:

- Finalizing work to provide more open systems or prepare suitable interfaces.
- A decision about the distribution or use through the networks of the specific tools developed for image treatment or enhancement.
- An assessment of the impact of prospective clients and adaptation of equipment to new needs.
- The essential use of broad-band networks if it is decided to provide remote service, not only for textual and descriptive information but also for digital images of documents. Otherwise, initial distribution of poorer quality images must be accepted.
- The search for another means of image service, bypassing human intervention to avoid initial loss of time, if images are consulted remotely. Solutions may include the use of jukebox or, even better, large-capacity magnetic disk sets to provide authentic online use of images. This would require another migration of images to a new carrier medium, which would be costly in terms of equipment and supplies and would require considerable time to complete.

Security Problems

Opening the Archivo's computerized system to remote consultation would require a strategy for addressing security problems, which have not been a source of concern so far. Much experience has been accumulated in the area of security, especially in important economic sectors such as banks, long-distance trading, industrial research, and in the military.

Problems of Intellectual Property and Document Ownership

The finding aids represent an intellectual creation that is the property of the state through the Ministry of Education and Culture. Distribution of this information, contained in the information and reference system, poses problems of intellectual property that are similar

in principle to those for any other information, whether or not fees are charged.

Furthermore, the document images themselves form part of a Spanish documentary heritage, the distribution of which requires careful decision making. Should free access be permitted? Limited access? Who is entitled to access?

There are several alternatives for distribution. Access may be free or a fee may be charged. A fee may initially be imposed, with subsquent access free of charge. Users may be free to further disseminate the information, or its use may be restricted to the authorized client. Use may be granted only to cultural institutions or individuals, or it may be extended to for-profit institutions.

The usual practice in Spanish archives has been to provide copies of documents on microfilm or paper to institutions and individuals. But complete series of documents are almost never provided, and in the exceptional cases when they are, a special agreement is required obligating the recipient to refrain from distributing copies of the documents. Should this practice be continued? Or should dissemination be expanded? Further restricted?

Regarding these questions, the experience of other centers is still limited. There is no record yet of a historical archive putting such a large volume of documents on the network. Whatever is done will affect future operation of the AGI and other archives. The decision should be carefully considered, ideally with the participation of all interested parties.

Management Problems

Certain management problems have already been noted. Still, the user-management system should be expanded to cover the following:

- New types of system access control
- New types of statistical data retrieval
- Accounting and billing systems, where applicable

There are other problems as well (minor, perhaps, but important in the practical operation of institutions), such as staffing and work schedules.

As in other aspects of society today, the "long-distance" or "virtual" reading room may soon become reality in archives, producing the long-distance researcher. The Archivo General de Indias could continue to progress in that direction and make a significant contribution.

CONCLUSIONS

Multidirectional Progress

Computerization of the Archivo General de Indias has been a highly ambitious operation in applying new technologies to archival functions. Its achievements in several areas are noteworthy:

- It has affected all areas of work in historical archives (management, information, consultation, and conservation), showing that it is possible, reasonable, and, ultimately, economically viable to undertake this type of integrated archival treatment system.
- It has generated eleven million digitized pages and incorporated all existing finding aids within the automated information and reference system.
- It has allowed experimentation at every point in the process, including functional analysis, monitoring several years of use in the Reading Room, and addressing the constant problems of obsolescence.
- It has allowed continued experimentation in the use of new archival technologies; the project was not abandoned at the first obstacle or under the pressure of constantly changing technologies.
- It has shown that the new technologies, properly used, offer powerful tools for attaining the Archivo's primary goals of conservation and dissemination.

Project Costs

Today, an analysis of the project's cost between 1986 and 1997 has undoubtedly lost some of its relevance, except as a historical footnote. In the world of new technology, prices of hardware, software, and media carriers are dropping rapidly, while prospects for new equipment are expanding. It is almost ridiculous to compare the price of a PC with a 486 processor in 1991 with that of a PC equipped with the latest generation processor in 1997. The price of hardware is halved about every 18 months.

That said, the total outlay has been high. During the project's main phase (1986-1992), the three institutions allocated, in equal shares, one billion pesetas (more than $US 6.6 million). The following two years required another 300 million ($US 2 million). This cost covered:

- Research, design, and development of the system
- Archival work involved in document preparation
- Data entry (textual databases and document digitization)
- Hardware and software
- Cost of storage media (particularly optical disks)

The largest item of expenditure has been personnel. The technical research and development group has always consisted of at least 10 people, and at times as many as 20.

The project also required experts in the archival treatment of documents. Tens of staff archivists from the Archivo General de Indias, the Archivo Histórico Nacional, and the Archivo General de Simancas have collaborated on the project, along with various groups of contract personnel with greater specialization and experience in such work. Finally, the project required data entry staff. They were needed first for the creation of the textual or descriptive database, with a group of seven operators working in Madrid for four years and another team of five in Seville. Then they were needed for the digitization of documents, with a team of as many as 32 operators in 1991 and 1992.

The cost of hardware and software has also been high but has decreased over time, accounting for an ever-smaller share of the whole. And the cost of storage media has been considerable.

With regard to costs, the following observations can be made:

- The important initial work of system research, design, and development drew on experience gained in other archives and in other projects. It represents a significant and, at least from the cultural standpoint, profitable "long-term investment," applicable beyond the AGI project.
- Since most of the archival work consisted of the usual activities in archives (organization and description of documents), it too extends beyond the digitization project per se, as does the entry of textual data.
- The cost of document digitization, accounting for a large percentage of total expenditures, was the most controversial item. Up to 32 employees worked together on a job that, thanks to technological progress, could now be performed with far fewer operators. But AGI would have had to wait six years to begin the work, and, as of today, would have had a scant five years of digital image use in the Reading Room. Nor would the AGI have enjoyed the other advantages of the system. If the decision had been delayed, would AGI now be certain that the time had come to do the work? Might staff not have decided to wait for new advances to allow more rapid and therefore less expensive work? It should be kept in mind that, owing to the type of documents involved and their state of conservation, automatic scanner feeding cannot be considered, which means that human intervention is inevitably a significant part of the entire process.

More information on actual costs is provided in Appendix 4.

Project Results

Project results can be summarized as follows:

- The integrated automation of all basic functions of the Archivo.
- The development of a unified data system containing all descriptive information following the retrospective conversion of finding aids.
- The replacement of consultation of original documents by a significant percentage of digitized documents, yielding benefits for conservation and access.
- Almost five years of continuous use of the system by researchers in the Reading Room and by the entire staff of the Archivo.
- Important benefits for the management and internal operation of the Archivo.
- Finally, a system model (hardware, software, and know-how) available for use in other archives.

More specifically, and in accordance with the initial aims of the project, the project has yielded the following benefits for the AGI and for research.

Conservation Benefits

Although digitization does not solve all the problems of document conservation, it has greatly reduced the risk of deterioration for about a third of the Archivo's original holdings. Over the past year, 31 percent of consultations at the AGI were done using the electronic document. Over the same period, paper copies made from the electronic document accounted for more than 38 percent of all paper reproductions delivered to researchers. It is reasonable, therefore, to state that, 31 percent of the risk of deterioration from document handling in the Reading Room and 38 percent of the risk from handling to make photocopies has been eliminated. Some of the original documents were at especially high risk because of their constant use. Between 1989 and 1992, some documents were consulted more than 40 times a year. Today, the most popular documents have all been digitized and are never handled. Appendix 1 provides further details.

Access Benefits

An analysis of consultation figures over the past few years shows that the researcher needs less time today to complete research thanks to the computerized system:

- Information access is more rapid.
- More information is available.
- Researchers complete more consultations per work session.
- The AGI provides more services to a broader range of users.
- Delivery of paper copies is much more rapid.

See Appendix 2 for more information.

Benefits for Internal Operation

Nearly all functions of the AGI are benefiting from the computerized system:

- Information service: response to the user is much more rapid and efficient for both on-site and correspondence consultations.
- Organization and description of holdings: the work of description and indexation, and the preparation of new catalogs, inventories, and indices has been expedited.
- Management: there have been improvements in tracking research requests, monitoring reading room use, moving documents, and preparing statistics.
- Reproduction service: requests for copies are processed and filled much more quickly.

Final Observations

In its efforts of the past ten years, AGI has consistently aimed to be practical and has shown that new technologies can offer powerful tools for fulfilling the major goals of conservation and dissemination.

- The AGI has converted eleven million pages and all existing descriptive data to digital format. That information is now available for the future. It may be transformed, adapted, and used in various ways, but it is there now.
- The use of information and experience accumulated in Spanish archives over a period of centuries also offers guarantees for the future. The system was not created in a vacuum but drew upon the entire historical memory of several centuries of archival treatment.
- Technologies change, and the system will have to be updated or replaced. But in the meantime, it will have made a significant contribution in service to a rich cultural heritage.

GENERAL BIBLIOGRAPHY

General References on the Archivo

González García, Pedro, ed. *Archivo General de Indias* (Madrid: Ministerio de Cultura, Lunwerg Editores, 1995).

González García, Pedro, ed. *Discovering the Americas: The Archive of the Indies* (Madrid: Ministerio de Educación y Cultura, Lunwerg Editores, 1997).

Peña y Cámara, José María de la. *Archivo General de Indias de Sevilla. Guía del Visitante* (Valencia: Dirección General de Archives y Bibliotecas, 1958).

The AGI's Origin and Ordinances

Ballesteros Beretta, Manuel. "Don Juan Bautista Muñoz, dos Facetas Scientíficas" [Don Juan Bautista Muñoz, Two Scientific Aspects]. *Revista de Indias*, no. 3 (1941): 5-37.

Ballesteros Beretta, Manuel. "Juan Bautista Muñoz: La Creación del Archivo de Indias" [Juan Bautista Muñoz: Establishment of the Archivo de Indias]. *Revista de Indias*, no. 4 (1941): 55-95.

Ballesteros Beretta, Manuel. "D. Juan Bautista Muñoz: La Historia del Nuevo Mundo" [D. Juan Bautista Muñoz: History of the New World]. *Revista de Indias*, no. 10 (1942): 589-660.

Gómez, Gómez, Margarita. *Fundación y Ordenanzas del Archivo General de Indias: Su Significación en la Política Archivística Española* [Foundation and Ordinances of the AGI: Its Role in Spanish Archival Policy]. (Seville: Universidad de Sevilla, 1993). (Microfiche).

Gómez, Gómez, Margarita. "El Archivo General de Indias. Génesis Histórica de sus Ordenanzas" [The AGI: Historical Origin of its Ordinances]. In *Ordenanzas del Archivo General de Indias* [Ordinances of the AGI]. Facsimile edition and preliminary studies (Seville, 1986), 53-120.

Muro Orejón, Antonio. "Juan Bautista Muñoz: Las Fuentes Bibliográficas de la Historia del Nuevo Mundo" [Juan Bautista Muñoz: Bibliographic Sources of New World History]. *Anuario de Estudios Americanos* 10 (1953): 265-337.

Romero Tallafigo, Manuel. "La Fundación del Archivo General de Indias" [The Founding of the AGI]. In *Archivo General de Indias*, ed.

Pedro González García (Madrid: Ministerio de Cultura, Lunwerg Editores, 1995), 33-52.

Serrera, Ramón María. "Manuel José de Ayala: Un Colaborador Decisivo en el Proyecto Historiográfico de Juan Bautista Muñoz" [Manuel José de Ayala: A Leading Collaborator in the Historiographic Project of Juan Bautista Muñoz]. In *Documentación y Archivos de la Colonización Española* Vol. 2 (Madrid: Ministerio de Cultura, 1980): 253-263.

Solano, Francisco de. "José de Gálvez. Fundador del Archivo de Indias" [José de Gálvez: Founder of the AGI]. In *Ordenanzas del Archivo General de Indias*. Facsimile edition and preliminary studies (Seville, 1986), 7-52.

General References on the Computerization Project

González García, Pedro. "Historical Documentation and Digital Conversion of Images at the Proyecto de Informatización of the Archivo General de Indias, Sevilla." *Microform Review* 18, no. 4 (1989): 217-21.

González García, Pedro. "Computerization Project for the Archivo General de Indias." In *International Conference: Archiving and Disseminating Historical Machine-readable Data* (Leiden, Netherlands, 1992).

González García, Pedro. "Novas Tecnologias no Arquivo Geral das Indias" [New Technologies at the AGI]. *Acervo, Revista do Arquivo Nacional de Rio de Janeiro* 7, nos. 1-2 (1994): 75-89.

González García, Pedro. "Proyecto de Informatización del Archivo General de Indias" [The Computerization Project of the AGI]. In *Actas del II Congreso Iberoamericano de Informática y Documentación. Mar del Plata (Argentina, 1988)* (Madrid: Centro Regional para la Enseñanza de la Informática, 1989), 409-12.

González García, Pedro. "Proyecto de Informatización del Archivo General de Indias" [The Computerization Project of the AGI]. In *Actas de las I Jornadas de Archivística de Euskadi, 1990. IRARGI* no. 3 (Vitoria, Spain: IRARGI, 1990): 261-81.

González García, Pedro. "El Archivo General de Indias y su Proyecto de Informatización. Nuevas Posibilidades Para la Investigación" [The AGI and its Computerization Project. New Research Possibilities]. *Cuadernos de Historia Moderna* no. 15 (1994): 231-49.

González García, Pedro. "Archivos y Nuevas Tecnologías: Perspectivas Para la Investigación" [Archives and New Technologies: Research Prospects]. In *Actas de las Jornadas Archivos e Investigación (Murcia, 1991)* (Murcia, 1996), 163-80.

Oddos, Jean-Paul. "La Numérisation des Archives Générales des Indes à Séville" [Digitization of the AGI in Seville]. *Bulletin de la Bibliothèque de France* 39, no. 4 (1994): 82-86.

Rütimann, Hans, and Stuart Lynn. *Computerization Project for the Archivo General de Indias, Seville, Spain. A Report to the Commission on Preservation and Access* (Washington, DC: Commission on Preservation and Access, 1992).

Vázquez de Parga, Margarita, and Pedro González García. "Changing Technologies in European Archives." *American Archivist* 55 (1992): 156-65.

On research conducted to develop image-enhancement tools to improve document legibility, see:

Bescós Ramón, Julián. "Image Processing Algorithms for Readability Enhancement of Old Manuscripts." In *Electronic Imaging 89*, vol. 1 (Pasadena, 1989): 392-97.

Bescós Ramón, Julián, Francisco Jaque, and Luis Montoto. "Reflectance and Optical Contrast of Old Manuscripts: Wavelength Dependence." *Scanning Imaging* Vol. 1,028 (Society of Photo-optical Instrumentation Engineers [SPIE], 1989): 258-62.

Bescós Ramón, Julián, Juan Pedro Secilla, and Juan Navarro. "Filtering and Compression of Old Manuscripts by Adaptive Processing Techniques." *Proceedings of the Society for Information Display International Symposium 1990* (Las Vegas: Society for Information Display, 1990): 384-87.

Bescós Ramón, Julián, Juan Navarro, and Carlos Ramón. "Mejora de Legibilidad de Documentos Antiguos Mediante Tratamiento Digital de Imágenes" [Enhancing Readability of Old Documents through Digital Image Treatment]. *IV Simposium Nacional de Reconocimiento de Formas y Análisis de Imágenes* [Proceedings of Fourth National Symposium on Form Recognition and Image Analysis] (Granada: Sociedad Española de Reconocimiento de Formas y Análisis de Imágenes, 1990): 51-58.

APPENDIX 1: Analysis of Conservation Results

Use of Computerized System

Since 1993, when the system became fully operative at the Archivo, its use has expanded steadily, as shown in Table 1, which compares the following factors for 1993-97:

- Digital image copies as a percentage of total paper reproductions delivered.
- On-screen document consultation as a percentage of total researcher work sessions.
- Both services are compared with the percentage of documents digitized vis-à-vis total existing documents.

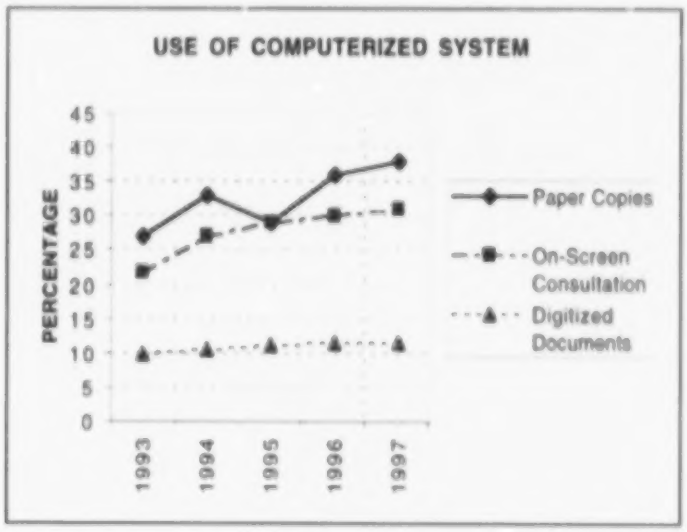

Table 1. *Percentages of consultation and paper reproduction directly through the computerized system as compared with total percentage of digitized documents.*

Data Provenance

Data used in the comparison were taken from two reports, "periodic movement of documents" and "services summary," derived from the system's user-management module:

- The first report provides, among other data, the number of visits or work sessions specifically dedicated to consulting original or digitized documents over a given period.

The total figures were:

	Original Consultation	*Digital Consultation*
1993	9,649	2,706
1994	7,711	2,893
1995	6,758	2,757
1996	6,474	2,771
1997 (5 months)	2,748	1,245

BEST COPY AVAILABLE

- Data on paper copies were obtained from another report, "service summary," derived from the user-management module. No comparison with microfilm use has been made because those figures were skewed by other circumstances, particularly the availability of AGI staff. Consequently, only the figures on paper copies—digitized images versus photocopies—have been compared.

	Photocopies	*Digital Images*
1993	143,201	53,906
1994	164,424	78,562
1995	131,842	54,364
1996	148,760	82,819
1997 (5 months)	64,774	41,041

Conclusion

Even though only 12 percent of all documents available at the AGI are in digital form, these documents account for about one-third of all service provided by the Archivo. This means that about one-third of the risk of document deterioration through consultations and paper reproduction has been eliminated.

APPENDIX 2: Analysis of Consultation Results

For researchers, the system's advantages are clear, although researchers themselves may be unaware of them. They may even be annoyed by surrendering the almost "fetishist" pleasure of touching the document. Still, the increased speed of work offsets that loss. We can unequivocally say that today's researchers require less time to do their work.

To illustrate this, we can look at consultation statistics in the Archivo General de Indias from 1989 to the end of 1996. Earlier years cannot be used since the system of recording consultation data then differed from the one employed after the first version of the user-management module was implemented in mid-1988. In addition, 1997 was eliminated because the analysis began before the year's end, and it was not possible to extrapolate data because the number and frequency of researchers' visits vary with the time of year and school schedule.

Increase in Number of Archivo Researchers

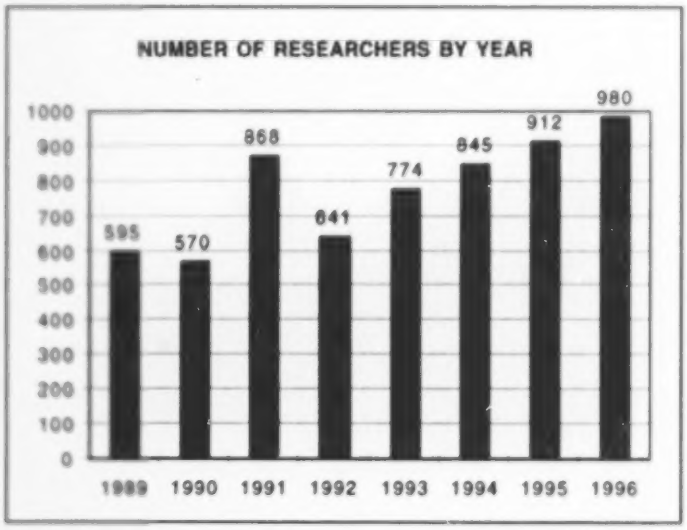

Table 2. *Researchers visiting the Archivo in recent years*

The most obvious feature of Table 2 is the overall steady increase since 1990 in the number of different researchers visiting the AGI. The "spike" in researchers in 1991 can be explained by the series of celebrations held in 1992. Many research projects were being completed in 1991 in preparation for the celebrations, while for much of 1992 it became very expensive to stay in Seville because of Expo 92.

Fewer Reading Room Work Sessions

In contrast, Table 3 shows a steady drop in the number of work sessions held each year by researchers at the AGI.

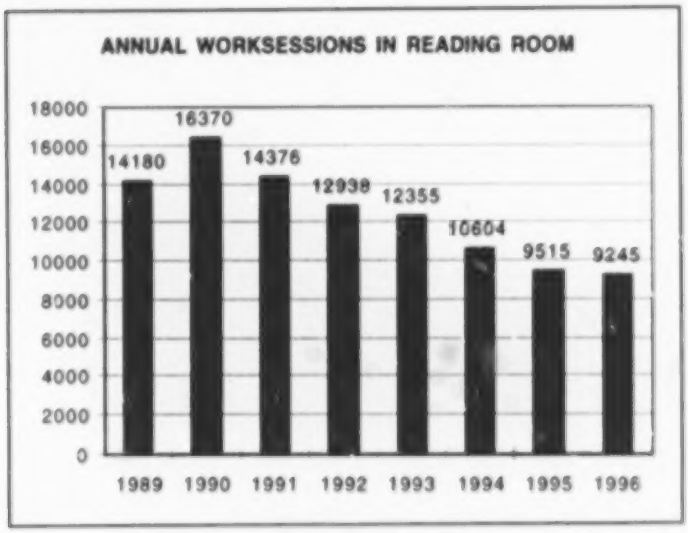

Table 3. *Total number of AGI researcher work sessions in recent years*

Increase in Number of Consultations/Day

This period also registered an increase in the average number of consultations by each researcher during a work session and a sharp drop in the number of sessions worked by each in the AGI. In the years before computerization, the average number of consultations per researcher and per day was just over two. As of 1993, however, it was more than three, with four-and-a-half consultations a day recorded in 1996 and 1997. See Table 4, based on the results of dividing the number of total consultations each year by the number of visits or work sessions.

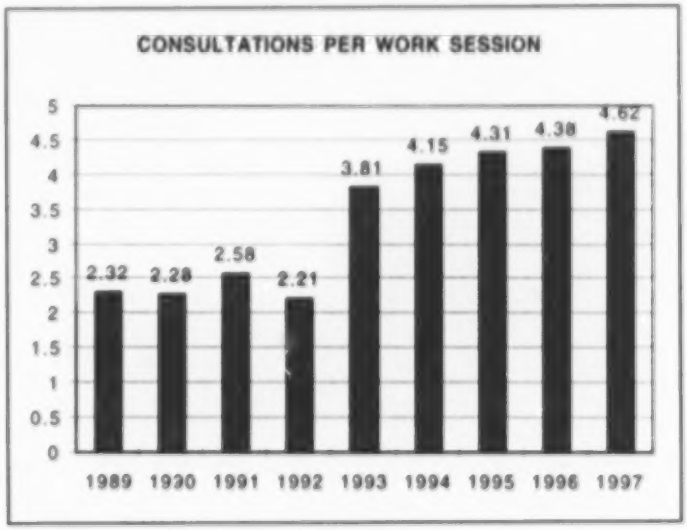

Table 4. *Average consultations per researcher per work session*

Fewer Work Sessions per Researcher

Other data show that the researcher spent much less time on average working in the AGI. Table 5 shows the progressive decline in recent years in the number of work sessions averaged by each researcher. The curve is extremely significant: work sessions per researcher dropped to fewer than half between 1992, the last year of manual consultation, and 1996, and no other reasons have been found to account for that decline. Hours of operation have remained the same and the number of reproductions delivered at the request of researchers has not risen much, yet the number of individual researchers visiting the Archivo has increased considerably.

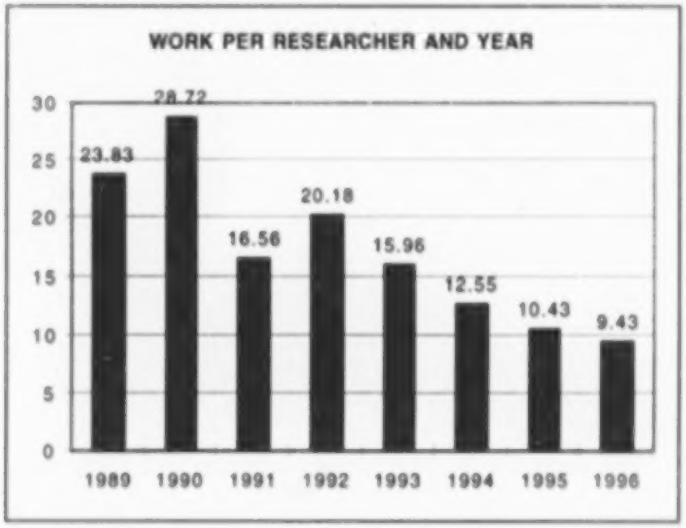

Table 5. *Average work sessions or visits per researcher and year*

Conclusions

The conclusions that can be drawn from the tables relate to the greater facility for research offered by the system: more individual researchers visit the Archivo, and yet there are fewer work sessions. Why? Because in each work session the researcher has access to more information and can consult more documents. These advantages can be analyzed in two areas: ease of finding documents and the ability to consult many on-screen.

- In the first instance, the information search is more rapid using the computerized system than using printed finding aids. The new system furnishes tools for easier, more rapid retrieval of data and much faster location of the original documents needed for consultation.
- The best tools are made available to researchers, and more information is provided to users today than was possible with the original finding aids, so information is not only easier to retrieve, but there is also more information available for locating documents.
- It takes only one or two minutes to obtain a requested document on-screen, while it can take several to retrieve the original from the stacks and deliver it to the Reading Room.
- Even if the researcher should decide to request a paper copy for subsequent consultation, service is still faster. While photocopies can be delivered to the AGI within one to several weeks, digital images

are usually delivered the day following a request, or even the same day if service fee payments are expedited.

In summary, there are two key advantages of the system for consultation. First, researchers need much less time for their work. Second, while more researchers visit the Archivo, they schedule fewer work sessions because they need less time for their research. Consequently, the Reading Room is less crowded.

APPENDIX 3: "Supplemental" Digitization of Documents from Other Archives—Analysis of Results

As noted in the section on selection, the body of digitized documents includes a sizable group of papers from the Archivo Histórico Nacional (AHN) and the Archivo General de Simancas (AGS). This is supplemental documentation concerning the administration of overseas territories, which, for one reason or another, failed to reach the Archivo General de Indias as ordered by King Carlos III[1].

There were two main reasons for the decision to digitize these supplemental holdings as part of the AGI project. The first relates to King Carlos' III original desire to bring together all materials "referring to the Indies." In this connection, the completed operation has been of significant value, difficult to quantify but tracing an interesting path for the "reconstruction of the archival heritage."[2]

The second aim of this operation has not yet been fully met. It was to incorporate other archival centers into the digitization process from the beginning, making two digital copies of the same papers (one for the AGI and the other for the archive holding the documents). As observed earlier in this report, the Ministry of Culture viewed computerization of the AGI as a pilot project for computerizing the other state historical archives.

Were these goals fulfilled? Only partially. The objective of supplementing AGI resources to make it, as King Carlos III had wished, an authentic "general archive" for papers about the Indies has been largely met by incorporating 2,189 new bundles in digital form.

But the aim of incorporating other major Spanish archival centers (particularly the AHN and the AGS) within the mainstream of new technologies has been only marginally accomplished (for example, the bundles digitized are not yet consulted in their own Reading Rooms).

What have been the results in terms of dissemination and conservation? Limited, because the holdings from other archives appear to be of less interest to AGI researchers. Table 6 shows that digitized documentation from the Archivo Histórico Nacional and Archivo General de Simancas is very seldom consulted in the AGI. It accounts for less than 10 percent of total consultations by means of digital image, even though it represents more than 40 percent (see Table 7) of the total documents digitized.

Those tables were based on the following data:

- The Archivo General de Indias currently consists of 43,209 bundles of documents, of which 3,210—or 7.43 percent of the entire Archivo—have been digitized.
- In addition, another 2,189 bundles from other centers (Archivo Histórico Nacional and Archivo General de Simancas) have been digitized (supplemental digitization).
- Adding the total figures for bundles (43,209 and 2,189 = 45,398) and comparing them with the bundles digitized (3,210 and 2,189 = 5,399), we can estimate that 11.89 percent of the total holdings of the three archives have been digitized. This 11.89 percent is serving 31.16 percent of the consultations in 1997, an excellent result.
- Of the total bundles digitized, those from the AGI account for 59 percent, while those from the AHN represent 31 percent and, from the AGS, 10 percent. In other words, the AGI has processed 59 percent as compared with 41 percent for the other two archives.
- However, the portion of documents from AHN and AGS consulted through the system in recent years has been only 6.04 percent.

[1] José María de la Peña y Cámara, "Cómo y porqué dejó de ser General el Archivo General de Indias. Cómo puede volver a serlo." (How and why did the Archivo General de Indias cease to be general. How it can become so again), in *Archivo Hispalense* 207 and 208 (1985): 21-40.

[2] See Pedro González García, "New Technology and the Reconstruction of the Archival Heritage," in *Proceedings of the XXX International Conference of the Round Table on Archives*, Thessaloniki, 1994. (Dordrecht, The Netherlands: International Council on Archives, 1998), 125-29.

This makes an important point regarding selection for digitization. Decisions about which of AGI's holdings should be digitized were based on an anlaysis of their use. This was not the case with documents from AHN or AGS. This drastically affects the results of system "use." What might the results have been for consultation and conservation if such material had not been incorporated into the digitization process?

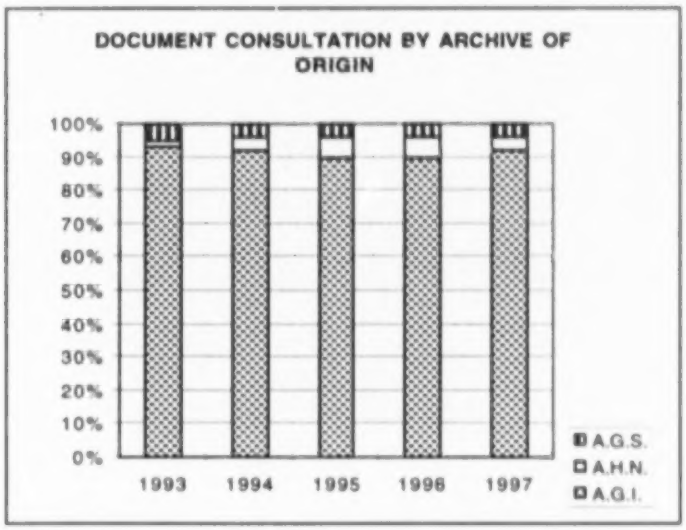

Table 6. *Consultations in the AGI Reading Room of digitized documents from the Archivo General de Indias, Archivo Histórico Nacional, and Archivo General de Simancas.*

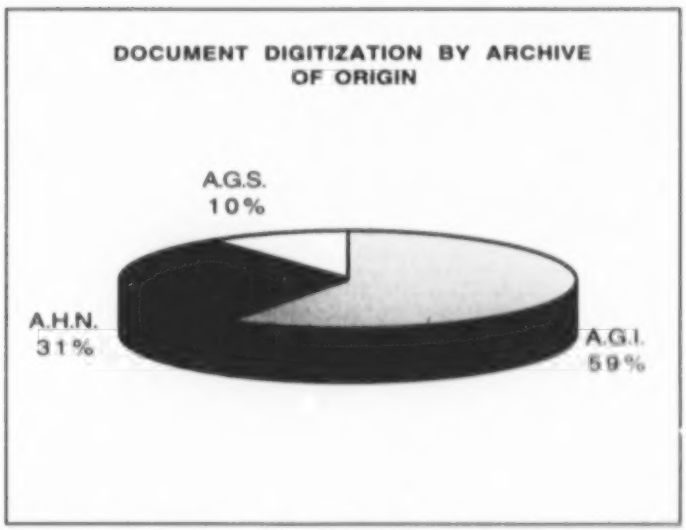

Table 7. *Percentage of digitized documents by archive of origin.*

APPENDIX 4: Figures for Costs and Production

The conclusion of this report stated that an analysis of costs during the main phases of the project has lost much of its relevance except as a historical footnote. But it is possible to give some data about the actual cost of implementing the system in an archive, to assess whether a project can be economically viable.

With this goal, we are going to analyze a typical case, including the acquisition of the necessary hardware and software and assuming production of a significant number of digital images (including document preparation, scanning, and storage).

The costs are estimated for Spain. Equipment and personnel costs vary greatly among countries.

1. The Documents

We assume that 250 bundles of old documents (about 11,300 files or 500,000 pages) will be scanned that pertain to two different series of the AGI: Confirmaciones de Encomiendas and Confirmaciones de Oficios. These are two series of the AGI that were digitized in the project's main phase. Their catalog is included in the information system and a paper copy also exists in the Reading Room. Preparation of the documents entails moderate difficulty.

2. The Work

Archival work

This involves the normal work of archiving: organizing, describing, indexing, placing files properly in folders, writing the call number, and so on. Description includes the cataloging of every file. Elements include title, dates, abstract, and call number. The indexing includes the extraction of the principal key words.

Textual data of the new descriptions are then entered in the information system, and a "digitization guide" is prepared. Repair or conservation of documents, sometimes necessary, is not assumed in this example.

Scanning

This involves the scanning, storage and creation of backup file, and quality control by automated means.

3. Staff and Production Estimates

Archival work

Three full-time staff per year (archival and historical background). It is presumed that an archives expert can prepare two bundles per week (45 weeks per year x 2 bundles per week x 3 persons = 270 bundles)

Scanning

Four full-time staff per year (scanner operator). The average production is 135,000 pages per person per year (600 pages per day x 5 days a week x 45 weeks a year). This is a conservative estimate. If there is no problem with hardware and personnel, productivity will be higher.

4. The Equipment

We assume that a full set of equipment is needed. It is composed of an Ethernet local area network, with:

- One server, with CD-R and DAT tape units, laser printer, Windows NT, DB2 or Oracle, and application software (ArchiGES and ArchiDOC).
- One scanning workstation with Windows NT and ArchiDOC for image capture.
- One textual data entry workstation (valid also for consultation), with Windows NT and the corresponding application software.
- One consultation workstation, with the same configuration.
- One user-management workstation, with a printer, windows NT, and the corresponding application software.

5. Project Duration

Preparation of documents is estimated to take one year, assuming a staff of three archivists. The level of difficulty in preparing various documents will influence the time and staffing required for preparation.

The scanning process requires four years if only one camera is used, and assuming one daily shift. If there are two daily shifts, the time required for scanning will be reduced to two years.

6. Costs

Personnel

Archives experts (3)	11,200,000 ptas.
Scanner operators (4)	10,500,000 ptas.

Hardware

Server (with laser printer, CD-R and DAT units)	1,740,000 ptas.
Digital camera	2,320,000 ptas.
Workstations (4)	1,850,000 ptas.
Media (disks, tapes, etc.)	100,000 ptas.
Basic software (Windows NT and Oracle)	700,000 ptas.
Application software (software, installation, training, hot line, etc.)	5,160,000 ptas.
TOTAL	33,570,000 ptas. ($US 223,800)

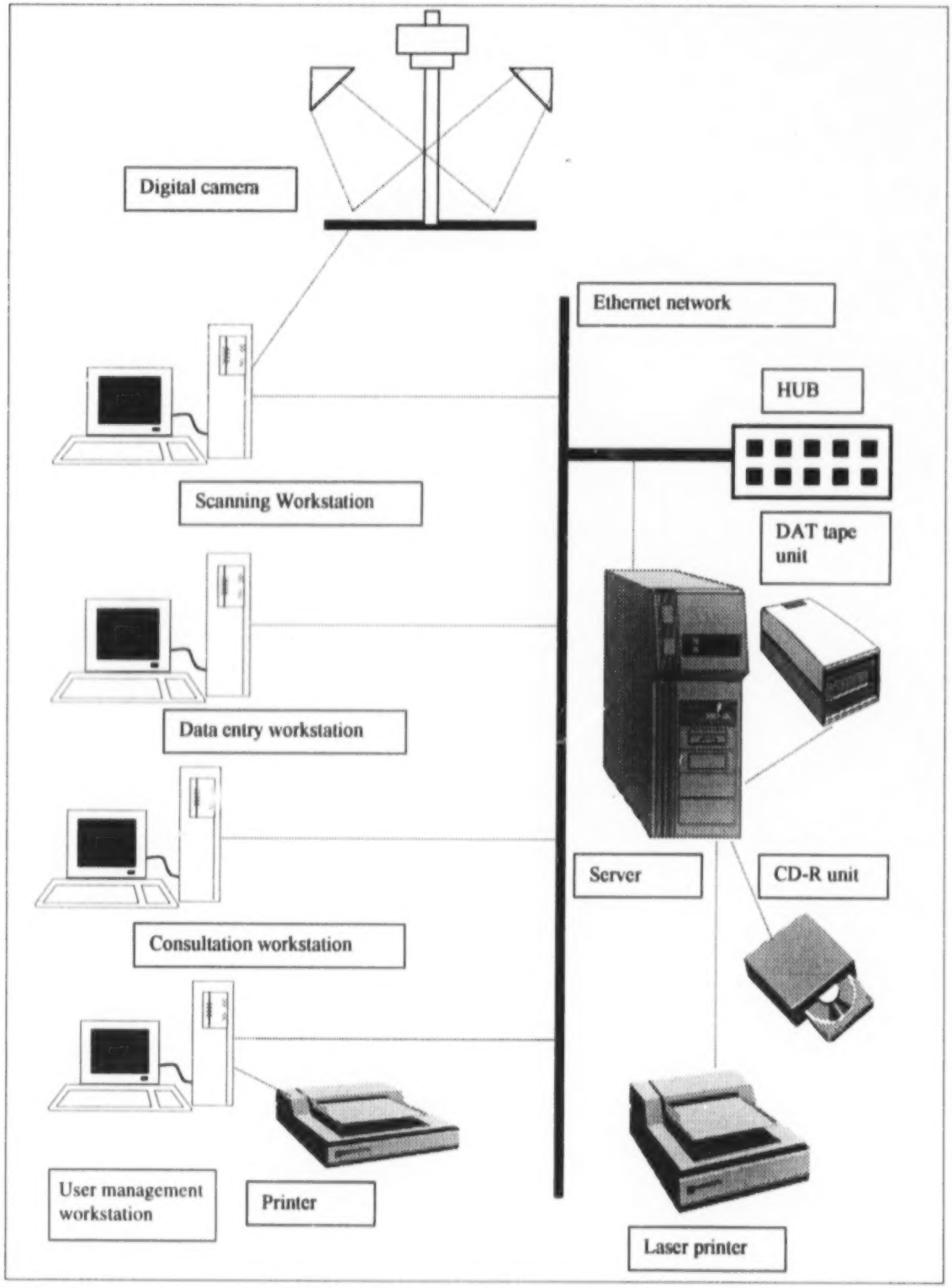

Figure 4. *Example of System Configuration*

BEST COPY AVAILABLE